

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 11:16:55 : Search time 1066.61 Seconds
(without alignments)
4101.602 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180
Sequence: 1 ggaagctcccttcctcaag.....tcataagaatggaatgc 180

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	em_estba:**
2:	em_esthum:**
3:	em_estlin:**
4:	em_estnu:**
5:	em_estcov:**
6:	em_estplo:**
7:	em_estro:**
8:	em_hlc:**
9:	gb_est1:**
10:	gb_est2:**
11:	gb_hlc:**
12:	gb_est3:**
13:	gb_est4:**
14:	gb_est5:**
15:	em_estfun:**
16:	em_estom:**
17:	em_gss_hum:**
18:	em_gss_inv:**
19:	em_gss_pln:**
20:	em_gss_vrc:**
21:	em_gss_vrc:**
22:	em_gss_mam:**
23:	em_gss_mus:**
24:	em_gss_pro:**
25:	em_gss_rod:**
26:	em_gss_phg:**
27:	em_gss_vrl:**
28:	gb_gss1:**
29:	gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	180	100.0	334	B0636767
2	180	100.0	335	B0636767
3	180	100.0	537	B0636767
4	180	100.0	553	B0636767

Result No.	Score	Query Match length	ID	Description
5	180	100.0	1115	B0636767
6	167.4	93.0	320	B0636767
7	165.2	91.8	220	B0636767
8	130.4	72.4	293	B0636767
9	130.4	72.4	325	B0636767
10	127.4	70.8	433	B0636767
11	120.8	67.1	285	B0636767
12	120.8	67.1	313	B0636767
13	120.8	67.1	318	B0636767
14	120.8	67.1	330	B0636767
15	120.8	67.1	352	B0636767
16	120.8	67.1	361	B0636767
17	120.8	67.1	420	B0636767
18	119.8	66.6	364	B0636767
19	119.4	66.3	316	B0636767
20	119.2	66.2	277	B0636767
21	117.6	65.3	395	B0636767
22	116.2	64.6	300	B0636767
23	114.4	63.6	277	B0636767
24	114.4	63.6	287	B0636767
25	112.8	62.7	300	B0636767
26	112.6	62.6	220	B0636767
27	112.6	62.6	229	B0636767
28	112.2	62.2	390	B0636767
29	111.2	61.8	277	B0636767
30	109.6	60.9	370	B0636767
31	109.2	60.7	369	B0636767
32	109.2	60.6	332	B0636767
33	108.2	60.1	350	B0636767
34	103.4	57.4	266	B0636767
35	100.8	56.0	242	B0636767
36	94	52.2	268	B0636767
37	90.4	50.2	282	B0636767
38	90.4	50.2	286	B0636767
39	88	48.9	214	B0636767
40	83.6	46.4	301	B0636767
41	83.4	46.3	146	B0636767
42	82	45.6	210	B0636767
43	81.2	45.1	835	B0636767
44	80.6	44.8	229	B0636767
45	69.8	38.8	402	B0636767

ALIGNMENTS

Result No.	Score	Query Match length	ID	Description
1	180	100.0	334	B0636767
2	180	100.0	335	B0636767
3	180	100.0	537	B0636767
4	180	100.0	553	B0636767

RESULT 1
LOCUS B0636767
DEFINITION he01b10.y1 Human Retina cDNA (Un-normalized, unamplified): hd/ne
ACCESSION B0636767
VERSION B0636767.1 GI:21761226
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS J.W., Bouffard, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A., Touchman
TITLE Expressed sequence tag analysis of human retina for the NEIRank Project: Retinidin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts
JOURNAL Mol. Vis. 8 (4), 2002
COMMENT Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel.: 301 402 3452
Fax: 301 496 0078
Email: graham@helix.nih.gov
Plate: 01 row: b column: 10

FEATURES
source
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1. .334
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="he01b10"
/issue_type="Retina"
/dev_stage="adult"
/lab_host="EMDH10B"
/clone_11b="Human Retina CDNA (Un-normalized, unamplified)
): hd/he"
/note="Organ: Eye; Vector: pSPOR1; Neural retina tissue was dissected from two 8 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPOR1 vector (Life Technologies) was constructed at Bioscience Biotechnology (Laurel MD) essentially following details of the SuperScript Plasmid System full instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor (5'-GGACTAGTCTAGATCGGAGCGGCCGCTT)5'-3'. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT
ORIGIN
94 a 86 c 88 g 66 t

Query Match
Best Local Similarity 100.0%; Score 180; DB 13; Length 334;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTTCAGAGATGCCATGTGACACATGTAGTCTCAACCTGT 60
|||||
Db 63 GGAAGCTCCCTTTTCAGAGATGCCATGTGACACATGTAGTCTCAACCTGT 122
|||||

QY 61 TCCAGATGTCCAACTGGTCTGGGACATGAGTGCATATAGATGATGTCAG 120
|||||
Db 123 TCCAGATGTCCAACTGGTCTGGGACATGAGTGCATATAGATGATGTCAG 182
|||||

QY 121 CTCCTGTGGCCCGGATAAACCAACAGACATCCAGATGAAAGATGCAATGC 180
|||||
Db 183 CTCCTGTGGCCCGGATAAACCAACAGACATCCAGATGAAAGATGCAATGC 242
|||||

RESULT 2
BM856263 355 bp mRNA linear EST 06-MAR-2002
LOCUS K-EST0139933 S14K402 Homo sapiens CDNA clone S14K402-51-E10 5',
DEFINITION mRNA sequence.
ACCESSION BM856263
VERSION BM856263.1 GI:19212662
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 355)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontiers Korean EST Project 2001
TITLE Unpublished
JOURNAL
COMMENT
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 51 row: E column: 10
High quality sequence stop: 355.
Location/Qualifiers

source
1. .355
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S14K402-51-E10"
/cell_line="K402"
/lab_host="Top10F"
/clone_11b="S14K402"
/note="Organ: Stomach; Vector: pTZ19RPI; Site:1: EcorI;
Site:2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deapped
with tabacco acid pyrophosphatase (TAP). The deapped
intact mRNA was ligated with DNA-RNA linker including Ecor
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcorI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT
ORIGIN
90 a 102 c 94 g 69 t

Query Match
Best Local Similarity 100.0%; Score 180; DB 12; Length 355;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTTCAGAGATGCCATGTGACACATGTAGTCTCAACCTGT 60
|||||
Db 119 GGAAGCTCCCTTTTCAGAGATGCCATGTGACACATGTAGTCTCAACCTGT 178
|||||

QY 61 TCCAGATGTCCAACTGGTCTGGGACATGAGTGCATATAGATGATGTCAG 120
|||||
Db 179 TCCAGATGTCCAACTGGTCTGGGACATGAGTGCATATAGATGATGTCAG 238
|||||

QY 121 CTCCTGTGGCCCGGATAAACCAACAGACATCCAGATGAAAGATGCAATGC 180
|||||
Db 239 CTCCTGTGGCCCGGATAAACCAACAGACATCCAGATGAAAGATGCAATGC 298
|||||

RESULT 3
BG207464 537 bp mRNA linear EST 21-APR-2001
LOCUS RST26945 Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.
DEFINITION BG207464
ACCESSION BG207464
VERSION BG207464.1 GI:13729151
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 537)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Kikka,A., Hess,J., Cottrien,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
TITLE
JOURNAL
MEDLINE 21227151
PUBMED 11329013
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 482.

FEATURES
source
Location/Qualifiers
1. 537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, In press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT
150 a 133 c 148 g 106 t

ORIGIN

Query Match 100.0%; Score 180; DB 10; Length 537;
Best Local Similarity 100.0%; Pred. No. 6e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGACACATGAGTCTCCAACTGT 60
|||||
270 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGACACATGAGTCTCCAACTGT 329
|||||
61 TCCGAGATGTCACACCTGCTGCGGCACTGATGGCTCACAATATGCAATGATGCCAG 120
|||||
Db 330 TCCGAGATGTCACACCTGCTGCGGCACTGATGGCTCACAATATGCAATGATGCCAG 389
|||||
Qy 121 CTCGCTTGGCCCGGATATAAACCAACAGGACATCCAGATCATGAAGATGGCAATGC 180
|||||
Db 390 CTCGCTTGGCCCGGATATAAACCAACAGGACATCCAGATCATGAAGATGGCAATGC 449
|||||

RESULT 4
AM960685 553 bp mRNA linear EST 01-JUN-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT

AM960685 553 bp mRNA linear EST 01-JUN-2000
EST372756 MAGE resequences, MAGF Homo sapiens cDNA, mRNA sequence.
AM960685
AM960685.1 GI:8150369
EST.
SOURCE
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 553)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspari, R., Gay, C., Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
Unpublished
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: jquackenbush@igf.org
Plate: 149
Seq primer: Reverse.

FEATURES
source
Location/Qualifiers
1. 553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="MAGE resequences, MAGF"
/note="Vector: pBluescriptsm"

BASE COUNT
156 a 153 c 137 g 107 t

ORIGIN

Query Match 100.0%; Score 180; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 6.1e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGACACATGAGTCTCCAACTGT 60
|||||
Db 123 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGACACATGAGTCTCCAACTGT 182
|||||
Qy 61 TCCGAGATGTCACACCTGCTGCGGCACTGATGGCTCACAATATGCAATGATGCCAG 120
|||||
Db 183 TCCGAGATGTCACACCTGCTGCGGCACTGATGGCTCACAATATGCAATGATGCCAG 242
|||||
Qy 121 CTCGCTTGGCCCGGATATAAACCAACAGGACATCCAGATCATGAAGATGGCAATGC 180
|||||
Db 243 CTCGCTTGGCCCGGATATAAACCAACAGGACATCCAGATCATGAAGATGGCAATGC 302
|||||

RESULT 5
BE543478 1115 bp mRNA linear EST 09-AUG-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT

BE543478 1115 bp mRNA linear EST 09-AUG-2000
601070792P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456839 5', mRNA sequence.
BE543478
BE543478.1 GI:9772123
EST.
SOURCE
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1115)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
Plate: LHM8445 row: h column: 24
High quality sequence stop: 385.

FEATURES
source
Location/Qualifiers
1. 1115
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="IMAGE:3456839"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_id="NIH_MGC_12"
/note="Organ: cervix; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.4 kb. Library prepared by Life Technologies."

BASE COUNT
480 a 300 c 247 g 88 t

ORIGIN

Query Match 100.0%; Score 180; DB 10; Length 1115;
Best Local Similarity 100.0%; Pred. No. 7.8e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGACACATGAGTCTCCAACTGT 60
|||||
Db 72 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGACACATGAGTCTCCAACTGT 131
|||||
Qy 61 TCCGAGATGTCACACCTGCTGCGGCACTGATGGCTCACAATATGCAATGATGCCAG 120
|||||
Db 132 TCCGAGATGTCACACCTGCTGCGGCACTGATGGCTCACAATATGCAATGATGCCAG 191
|||||
Qy 121 CTCGCTTGGCCCGGATATAAACCAACAGGACATCCAGATCATGAAGATGGCAATGC 180
|||||
Db 192 CTCGCTTGGCCCGGATATAAACCAACAGGACATCCAGATCATGAAGATGGCAATGC 251
|||||

RESULT 6
AA534438/c 321 bp mRNA linear EST 21-AUG-1997
LOCUS n18004.x51 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:926191 3'
DEFINITION similar to SW:PE60_PIG P37109 PEPTIDE PEC-60 PRECURSOR. ;, mRNA
sequence.
ACCESSION AA534438
VERSION AA534438.1 GI:2278691
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 321)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrr/image/image.html
Insert Length: 417 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham.
FEATURES
source
1..321
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:926191"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dt) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p773 vector. Library went through one round of
normalization."
BASE COUNT 62 a 83 c 81 g 95 t
ORIGIN
Query Match 93.0%; Score 167.4; DB 9; Length 321;
Best Local Similarity 98.9%; Pred. No. 4e-41;
Matches 179; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCACCTGT 60
|||||
279 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCACCTGT 220
|||||
61 TCCAGATGTCCAACTGTGCTGGCGGCTGATGGGCTCACAATATGCAATG 119
|||||
219 TCCAGATGTCCAACTGTGCTGGCGGCTGATGGGCTCACAATATGCAATG 160
|||||
120 GCTCTGCTGGCGGCGGATATAAACCAAGACAGATCCAGATCATGGAATGGCAAT 179
|||||
Db 159 GCTCTGCTGGCGGCGGATATAAACCAAGACAGATCCAGATCATGGAATGGCAAT 100
|||||
QY 180 C 180
1
Db 99 C 99

RESULT 7

AT1732343/c 220 bp mRNA linear EST 13-DEC-1999
LOCUS n18004.x5 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:926191 3'
DEFINITION similar to TR:060575 060575 GASTROINTESTINAL PEPTIDE. ;, mRNA
sequence.
ACCESSION AT1732343
VERSION AT1732343.1 GI:5053456
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 220)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrr/image/image.html
This read is a RESUBMISSION of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Trace considered overall poor quality
Insert Length: 417 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES
source
1..220
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:926191"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dt) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p773 vector. Library went through one round of
normalization."
BASE COUNT 49 a 55 c 57 g 59 t
ORIGIN
Query Match 91.8%; Score 165.2; DB 9; Length 220;
Best Local Similarity 95.5%; Pred. No. 1.7e-40;
Matches 170; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCACCTGT 60
|||||
Db 178 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCACCTGT 119
|||||
61 TCCAGATGTCCAACTGTGCTGGCGGCTGATGGGCTCACAATATGCAATG 120
|||||
Db 118 TCCAGATGTCCAACTGTGCTGGCGGCTGATGGGCTCACAATATGCAATG 59
|||||
QY 121 CTCTGCTGGCGGCGGATATAAACCAAGACAGATCCAGATCATGGAATGGCAAT 178
|||||
Db 58 CTCTGCTGGCGGCGGATATAAACCAAGACAGATCCAGATCATGGAATGGCAAT 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 10:26:50 ; Search time 783.841 Seconds

(without alignments)
9394.424 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180

Sequence: 1 ggaagctcccttcctcaag.....tcataaagatgcaaatgc 180

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1.*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rdd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	100.0	180	6	AR166137
2	180	100.0	180	6	E13090
3	180	100.0	386	9	AF048700
4	180	100.0	388	6	AR028526
5	180	100.0	388	6	AR075817
6	180	100.0	388	6	BD007521
7	180	100.0	398	6	AR166138
8	180	100.0	398	6	E13091
9	180	100.0	402	6	AY014851
10	127.2	70.7	364	4	S46866
11	127.2	70.7	364	4	SSPEC60
12	120.8	67.1	515	10	MMMPGCC60
13	117.4	65.2	163396	9	AL161445
14	117.4	65.2	176584	2	AC027812
15	117.4	65.2	183099	2	AC034144
16	117.4	65.2	188540	2	AC011108
17	81.2	45.1	192093	2	AC121205
18	81.2	45.1	230493	10	AL837521
19	81.2	45.1	242175	2	AC096463
20	41.4	23.0	168	6	I08731
21	41.4	23.0	182	6	AI1350
22	41.4	23.0	182	6	AI1352
23	41.4	23.0	182	6	A24905
24	41.4	23.0	182	12	SYNPSTIA
25	41.4	23.0	186	6	A24907
26	41.4	23.0	482	6	E03479
27	39.8	22.1	5635	12	ASVPSKAN8
28	38.2	21.2	368	6	AX332609
29	38.2	21.2	368	6	AX333248
30	38.2	21.2	368	6	AX411203
31	38.2	21.2	368	6	BD160949
32	38.2	21.2	368	9	HSPST1
33	38.2	21.2	369	10	MMPIP12
34	36.6	20.3	179	6	E01725
35	36.6	20.3	179	6	E02456
36	36.6	20.3	204	12	SYNPSTIB
37	36.6	20.3	234	6	E01819
38	36.6	20.3	234	6	E03120
39	36.6	20.3	302	6	AR059723
40	36.6	20.3	302	6	E08411
41	36.6	20.3	341	6	E01574
42	36.6	20.3	341	12	SYNCPST
43	36.6	20.3	362	9	BC025790
44	36.6	20.3	432	6	AR270687
45	36.6	20.3	432	9	HMPSTI

ALIGNMENTS

RESULT 1
LOCUS AR166137
DEFINITION Sequence 1 from patent US 6280968.
ACCESSION AR166137
VERSION AR166137.1 GI:16241312
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 180)
AUTHORS Kato,S., Yamaguchi,T., Sekine,S. and Kamata,K.
TITLE Human PEC-60-like protein and DNA encoding the same
JOURNAL Patent: US 6280968-A 1 28-AUG-2001;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

source	1. .180	/organism="unknown"	
BASE COUNT	53 a	49 c	40 g
ORIGIN			38 t
Query Match	100.0%;	Score 180;	DB 6;
Best Local Similarity	100.0%;	Pred. No. 2.6e-47;	Length 180;
Matches 180;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;
QY	1	GGAAGAGTCCCTTCTCTCAAGAAATGCCCATCTGTGTGAACACATGTAGAGATCTCCAACTGT	60
DB	1	GGAAGAGTCCCTTCTCTCAAGAAATGCCCATCTGTGTGAACACATGTAGAGATCTCCAACTGT	60
QY	61	TCCCAAGATGTCCAAACCTGCTGTCGGGCACTGATGAGTGGCTCACATATACGAATGAATGCCAG	120
DB	61	TCCCAAGATGTCCAAACCTGCTGTCGGGCACTGATGAGTGGCTCACATATACGAATGAATGCCAG	120
QY	121	CTGCGCTTGGCCCGGATATAAACCAACAGACATCCAGATCATGAAGATGGCAATGC	180
DB	121	CTGCGCTTGGCCCGGATATAAACCAACAGACATCCAGATCATGAAGATGGCAATGC	180
RESULT 2			
LOCUS	E13090	180 bp	DNA
DEFINITION	Human an cDNA encoding a PEC-60-like protein.	linear	PAT 27-APR-1998
ACCESSION	E13090		
VERSION	E13090.1	GI:3251902	
KEYWORDS	JP 1997124698-A/1.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1. (bases 1 to 180)		
AUTHORS	Kato,M., Yamaguchi,T., Sekine,S. and Kamata,M.		
TITLE	HUMAN PEC-60-LIKE PROTEIN AND DNA CARABE OF CODING THE SAME		
JOURNAL	Patent: JP 1997124698-A 1 13-MAY-1997;		
COMMENT	SAGAMI CHEM RES CENTER		
	OS Homo sapiens (human)		
	PN JP 1997124698-A/1		
	PD 13-MAY-1997		
	PF 27-OCT-1995 JP 1995280272		
	PI KATO MASASHI, YAMAGUCHI TOMOKO, SEKINE SHINGO, PI KAWATA		
	MITSUHIISA		
	PC C07K14/575,C07H21/04,C12N15/09//A61K38/22,A61K38/22,A61K38/22,		
	PC C12P1/02,		
	PC (C12P21/02,C12R1.19);		
	CC strandedness: Double;		
	CC topology: linear;		
	EH key		
	FT source		
	FT 1. .180		
	FT /organism='Homo sapiens'		
	FT /tissue-type='stomach cancer'		
	FT mat_peptide		
FEATURES			
source	1. .180		
	Location/Qualifiers		
	/product='PEC-60-like protein'.		
	/organism='Homo sapiens'		
	/mol_type='genomic DNA'		
	/db_xref='taxon:9606'		
BASE COUNT	53 a	49 c	40 g
ORIGIN			38 t
Query Match	100.0%;	Score 180;	DB 6;
Best Local Similarity	100.0%;	Pred. No. 2.6e-47;	Length 180;
Matches 180;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;
QY	1	GGAAGAGTCCCTTCTCTCAAGAAATGCCCATCTGTGTGAACACATGTAGAGATCTCCAACTGT	60
DB	1	GGAAGAGTCCCTTCTCTCAAGAAATGCCCATCTGTGTGAACACATGTAGAGATCTCCAACTGT	60
QY	61	TCCCAAGATGTCCAAACCTGCTGTCGGGCACTGATGAGTGGCTCACATATACGAATGAATGCCAG	120
DB	61	TCCCAAGATGTCCAAACCTGCTGTCGGGCACTGATGAGTGGCTCACATATACGAATGAATGCCAG	120
QY	121	CTGCGCTTGGCCCGGATATAAACCAACAGACATCCAGATCATGAAGATGGCAATGC	180
DB	121	CTGCGCTTGGCCCGGATATAAACCAACAGACATCCAGATCATGAAGATGGCAATGC	180

[illegible]

Unclassified.
1 (bases 1 to 388)
REFERENCE
AUTHORS Bandman,O., Goli,S.K. and Murry,L.E.
TITLE Tumor-associated kazal inhibitor-like polypeptides and encoding polynucleotides
JOURNAL Patent: US 5858210-A 2 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..388
/organism="unknown"
BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN

Query Match 100.0%; Score 180; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACAGTGGTGTCCCACTGT 60
DB 111 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACAGTGGTGTCCCACTGT 170
DB 61 TCCGAGATGCCCACTGTGTCGGCAGCTGATGGGCTCACAATATAGCAATGATGCCAG 120
DB 171 TCCGAGATGCCCACTGTGTCGGCAGCTGATGGGCTCACAATATAGCAATGATGCCAG 230
QY 121 CTCTGCTGGCCCGGATATAAACCAACAGACATCCAGATCATATAAAGATGGCAATGC 180
DB 231 CTCTGCTGGCCCGGATATAAACCAACAGACATCCAGATCATATAAAGATGGCAATGC 290

RESULT 5
AR075817 388 bp DNA linear PAT 30-AUG-2000
LOCUS AR075817
DEFINITION Sequence 2 from patent US 5958699.
ACCESSION AR075817
VERSION AR075817.1 GI:10002563
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 388)
AUTHORS Bandman,O., Goli,S.K. and Murry,L.E.
TITLE Tumor-associated kazal inhibitor
JOURNAL Patent: US 5958699-A 2 28-SEP-1999;
FEATURES Location/Qualifiers
source 1..388
/organism="unknown"
BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN

Query Match 100.0%; Score 180; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACAGTGGTGTCCCACTGT 60
DB 111 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACAGTGGTGTCCCACTGT 170
QY 61 TCCGAGATGCCCACTGTGTCGGCAGCTGATGGGCTCACAATATAGCAATGATGCCAG 120
DB 171 TCCGAGATGCCCACTGTGTCGGCAGCTGATGGGCTCACAATATAGCAATGATGCCAG 230
QY 121 CTCTGCTGGCCCGGATATAAACCAACAGACATCCAGATCATATAAAGATGGCAATGC 180
DB 231 CTCTGCTGGCCCGGATATAAACCAACAGACATCCAGATCATATAAAGATGGCAATGC 290

RESULT 6
BD007521 388 bp DNA linear PAT 31-JAN-2002
LOCUS BD007521
DEFINITION Tumor-associated KAZAL type inhibitor.
ACCESSION BD007521
VERSION BD007521.1 GI:18635894
KEYWORDS JP 2001503629-A/1.

SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 388)
AUTHORS Bandman,O., Goli,S.K. and Murry,L.E.
TITLE Tumor-associated KAZAL type inhibitor
JOURNAL Patent: JP 2001503629-A 1 21-MAR-2001;
COMMENT INCTE PHARMACEUTICALS INC
OS Unidentified
PN JP 2001503629-A/1
PD 21-MAR-2001
PF 05-NOV-1997 JP 1998521770
PI 06-NOV-1996 US 08/744670
PC OLGA BANDMAN,SURYA K GOLI,LYNN E MURRY
PC C12N15/09,A61K38/55,A61K39/395,A61P1/00,A61P35/00,C07K14/81,
PC C07K16/38,
PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12Q1/68,G01N33/574, PC
C12N15/00,C12N5/00,
PC A61K37/64
CC
FH Key Location/Qualifiers
FT source 1..388
FT /organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN

Query Match 100.0%; Score 180; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACAGTGGTGTCCCACTGT 60
DB 111 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACAGTGGTGTCCCACTGT 170
QY 61 TCCGAGATGCCCACTGTGTCGGCAGCTGATGGGCTCACAATATAGCAATGATGCCAG 120
DB 171 TCCGAGATGCCCACTGTGTCGGCAGCTGATGGGCTCACAATATAGCAATGATGCCAG 230
QY 121 CTCTGCTGGCCCGGATATAAACCAACAGACATCCAGATCATATAAAGATGGCAATGC 180
DB 231 CTCTGCTGGCCCGGATATAAACCAACAGACATCCAGATCATATAAAGATGGCAATGC 290

RESULT 7
AR166138 398 bp DNA linear PAT 17-OCT-2001
LOCUS AR166138
DEFINITION Sequence 2 from patent US 6280968.
ACCESSION AR166138
VERSION AR166138.1 GI:16241313
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 398)
AUTHORS Kato,S., Yamaguchi,T., Sekine,S. and Kamata,K.
TITLE Human PEC-60-like protein and DNA encoding the same
JOURNAL Patent: US 6280968-A 2 28-AUG-2001;
FEATURES Location/Qualifiers
source 1..398
/organism="unknown"
BASE COUNT 107 a 110 c 105 g 76 t
ORIGIN

Query Match 100.0%; Score 180; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACAGTGGTGTCCCACTGT 60

Db 121 GGAAGGTCCTTCTCAAGATGCGCATCTGTGAACATGTAAGTCTCCAACTGT 180
|||||
QY 121 TCCAGATGTCAACACTGCTCTGGGACATGATGGCTCAATATGAAATGATGCAG 120
|||||
Db 181 TCCAGATGTCAACACTGCTCTGGGACATGATGGCTCAATATGAAATGATGCAG 240
|||||
QY 121 CTCGCTTGGCCCGGATTAACCAACAGACATCCAGATCATGAAGATGGCAATGC 180
|||||
Db 241 CTCGCTTGGCCCGGATTAACCAACAGACATCCAGATCATGAAGATGGCAATGC 300
|||||
RESULT 8
LOCUS E13091 398 bp DNA linear PAT 27-APR-1998
DEFINITION Human cDNA encoding a PEC-60-like protein.
ACCESSION E13091
VERSION E13091.1 GI:3251903
KEYWORDS JP 1997124698-A/2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 398)
AUTHORS Kato,M., Yamaguchi,T., Sekine,S. and Kamata,M.
TITLE HUMAN PEC-60LIKE PROTEIN AND DNA CAPABLE OF CODING THE SAME
JOURNAL Patent: JP 1997124698-A 2 13-MAY-1997;
SAGAMI CHEM RES CENTER
COMMENT OS Homo sapiens (human)
PN JP 1997124698-A/2
PD 13-MAY-1997
PF 27-OCT-1995 JP 1995280272
PI KATO MASASHI, YAMAGUCHI TOMOKO, SEKINE SHINGO, PI KAMATA
MITSUHIISA
PC C07K14.5/75, C07H21/04, C12N15/09//A61K38/22, A61K38/22,
C12P21/02,
PC (C12P21/02, C12R1:19);
CC strandness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..398
FT mat_peptide 43..120 /product='PEC-60-like protein precursor' FT
FT 43..303 /product='PEC-60-like protein' FT CDS
FEATURES
source Location/Qualifiers
1..398
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 107 a 110 c 105 g 76 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGAAGGTCCTTCTCAAGATGCGCATCTGTGAACATGTAAGTCTCCAACTGT 60
|||||
Db 121 GGAAGGTCCTTCTCAAGATGCGCATCTGTGAACATGTAAGTCTCCAACTGT 180
|||||
QY 61 TCCAGATGTCAACACTGCTCTGGGACATGATGGCTCAATATGAAATGATGCAG 120
|||||
Db 181 TCCAGATGTCAACACTGCTCTGGGACATGATGGCTCAATATGAAATGATGCAG 240
|||||
QY 121 CTCGCTTGGCCCGGATTAACCAACAGACATCCAGATCATGAAGATGGCAATGC 180
|||||
Db 241 CTCGCTTGGCCCGGATTAACCAACAGACATCCAGATCATGAAGATGGCAATGC 300
|||||

RESULT 9
LOCUS AX014851 402 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 41 from Patent WO953040.
ACCESSION AX014851
VERSION AX014851.1 GI:10041118
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from ovarian tumour tissue
JOURNAL Patent: WO 953040-A 41 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
source Location/Qualifiers
1..402
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 108 a 111 c 107 g 76 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGAAGGTCCTTCTCAAGATGCGCATCTGTGAACATGTAAGTCTCCAACTGT 60
|||||
Db 124 GGAAGGTCCTTCTCAAGATGCGCATCTGTGAACATGTAAGTCTCCAACTGT 183
|||||
QY 61 TCCAGATGTCAACACTGCTCTGGGACATGATGGCTCAATATGAAATGATGCAG 120
|||||
Db 184 TCCAGATGTCAACACTGCTCTGGGACATGATGGCTCAATATGAAATGATGCAG 243
|||||
QY 121 CTCGCTTGGCCCGGATTAACCAACAGACATCCAGATCATGAAGATGGCAATGC 180
|||||
Db 244 CTCGCTTGGCCCGGATTAACCAACAGACATCCAGATCATGAAGATGGCAATGC 303
|||||
RESULT 10
LOCUS S46866 364 bp mRNA linear MAM 05-AUG-1999
DEFINITION PEC-60-gastrointestinal peptide [swine, duodenum, mRNA, 364 nt].
ACCESSION S46866
VERSION S46866.1 GI:257597
KEYWORDS
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 364)
AUTHORS Metts,M., Cintra,A., Solfrini,V., Ernors,P., Bortolotti,F.,
Morasutti,D.G., Ostenson,C.G., Efendic,S., Agerberth,B., Mutt,V.
et.al.
TITLE Molecular cloning of PEC-60 and expression of its mRNA and peptide
JOURNAL In the gastrointestinal tract and immune system
MEDLINE J. Biol. Chem. 267 (28), 19829-19832 (1992)
PUBMED 1400298
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI g15614] from the original journal article.
This sequence comes from Fig. 1.
FEATURES
source Location/Qualifiers
1..364
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"

Db 302 CTTGCTGACCCGGATGAAACCATGACATTCATCATGAGAGTGCATGC 361
 RESULT 13
 AL161445
 LOCUS
 DEFINITION AL161445 163396 bp DNA linear PRI 04-DEC-2001
 Human DNA sequence from clone RP11-326F20 on chromosome 9 contains
 the GGT1 gene for glycoprotein alpha-galactosyl transferase 1, the
 SPINK4 gene for Kazal type 4 serine protease inhibitor, the 3' end
 of the BAG1 gene for BCL2-associated athanogene and a Cpg island,
 complete sequence.
 AL161445
 VERSION AL161445.10 GI:11137678
 KEYWORDS HTG; athanogene; BAG1; BCL2; Cpg island; galactosyltransferase;
 GGT1; glycoprotein; Kazal; serine protease inhibitor; SPINK4.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 163396)
 AUTHORS Ramsay, H.
 JOURNAL Direct Submission
 Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Nov 10, 2000 this sequence version replaced gi:11071610.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr9
 This sequence is the entire insert of clone RP11-326F20 The true
 left end of clone RP11-344B24 is at 100366 in this sequence. The
 true right end of clone RP11-54K16 is at 54767 in this sequence.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP11-326F20 is from
 the library RPCT-11.2 constructed by the group of Pieter de Jong.
 For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6.
 FEATURES
 source
 Location/Qualifiers
 1. 163396
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-326F20"
 /clone_lib="RPCT-11.2"
 5. 237
 repeat_region
 /note="AluY repeat: matches 77. .309 of consensus"
 239. 314
 repeat_region
 /note="L1M4 repeat: matches 2591. .2666 of consensus"
 331. 438
 repeat_region
 /note="MER44A repeat: matches 225. .331 of consensus"
 440. 555
 repeat_region
 /note="AluSq/x repeat: matches 11. .126 of consensus"
 566. 708
 repeat_region

repeat_region
 /note="AluO/FAM repeat: matches 158. .298 of consensus"
 795. 1193
 repeat_region
 /note="L1M4 repeat: matches 3657. .4070 of consensus"
 1194. 1491
 repeat_region
 /note="AluSp repeat: matches 1. .296 of consensus"
 1492. 1680
 repeat_region
 /note="L1M4 repeat: matches 4070. .4265 of consensus"
 1681. 1981
 repeat_region
 /note="AluSx repeat: matches 1. .301 of consensus"
 1982. 2181
 repeat_region
 /note="L1M4 repeat: matches 4265. .4471 of consensus"
 2182. 2486
 repeat_region
 /note="Aluub repeat: matches 1. .304 of consensus"
 2487. 2783
 repeat_region
 /note="L1M4 repeat: matches 4471. .4701 of consensus"
 3046. 3192
 repeat_region
 /note="AluSq/x repeat: matches 1. .147 of consensus"
 3196. 3247
 repeat_region
 /note="MIR repeat: matches 17. .66 of consensus"
 4058. 4366
 repeat_region
 /note="MER41B repeat: matches 150. .635 of consensus"
 4567. 4862
 repeat_region
 /note="AluSq repeat: matches 1. .300 of consensus"
 4863. 5008
 repeat_region
 /note="MER41B repeat: matches 1. .150 of consensus"
 6433. 6490
 repeat_region
 /note="MIR repeat: matches 81. .138 of consensus"
 7523. 7596
 repeat_region
 /note="L1P1 repeat: matches 5753. .5825 of consensus"
 8418. 8461
 repeat_region
 /note="L1M4 repeat: matches 6132. .6184 of consensus"
 8462. 8769
 repeat_region
 /note="AluSx repeat: matches 1. .304 of consensus"
 8770. 8783
 repeat_region
 /note="L1M4 repeat: matches 6145. .6132 of consensus"
 8784. 9085
 repeat_region
 /note="AluSp repeat: matches 1. .302 of consensus"
 9086. 9106
 repeat_region
 /note="L1M4 repeat: matches 6122. .6146 of consensus"
 9152. 9751
 repeat_region
 /note="L1M3 repeat: matches 5512. .6125 of consensus"
 10464. 10534
 repeat_region
 /note="MIR repeat: matches 61. .130 of consensus"
 11734. 11790
 repeat_region
 /note="L1 repeat: matches 2688. .2742 of consensus"
 11747. 11869
 repeat_region
 /note="MIR repeat: matches 132. .252 of consensus"
 11874. 12110
 repeat_region
 /note="FLAMC repeat: matches 1. .134 of consensus"
 12121. 12219
 repeat_region
 /note="MIR repeat: matches 35. .142 of consensus"
 12243. 12346
 repeat_region
 /note="52 copies 2 mer tg 71% conserved"
 12254. 12343
 repeat_region
 /note="9 copies 10 mer ggtgtgtgt 72% conserved"
 complement(12461. .12790)
 /note="match: GSS: Em:AO210501"
 complement(12571. .12678)
 /note="match: GSS: Em:AO177335"
 12602. 13170
 repeat_region
 /note="MER21B repeat: matches 232. .794 of consensus"
 13171. 13289
 repeat_region
 /note="FLAM A repeat: matches 1. .122 of consensus"
 13290. 13517
 repeat_region
 /note="MER21B repeat: matches 4. .232 of consensus"
 14033. 14159
 repeat_region
 /note="MIR repeat: matches 112. .257 of consensus"
 complement(14283. .70891)
 /gene="GGTA1"
 mRNA
 complement(join(14283. 17232,17420. 17524,19637. 19759,
 24065. 24252,38835. 39070,70404. 70891))
 /gene="GGTA1"
 /product="ba326F20.1 (glycoprotein, alpha-galactosyl

transferase 1)"
/note="glycoprotein, alpha-galactosyl transferase 1
match: cDNAs: Em:J03880 Em:D00314 Em:E02227 Em:M13566
Em:X14085 Em:A23697 Em:X14558 Em:M22921 Em:U10473
Em:E02228 Em:X13223 Em:D29805 Em:M13214 Em:X55415
Em:U19890 Em:Y12510 Em:M13701 Em:U19889 Em:AF038660
Em:U10474 Em:AB024434 Em:AB019541 Em:U10472 Em:AF142670
match: ESTs: Em:BF124086 Em:AA667118 Em:AA769459
Em:AA476930 Em:AI192425 Em:AI572046 Em:AI193377
Em:AA830559 Em:AA889902 Em:AI028425 Em:AW362033
Em:AI149203 Em:AI565973"
/evidence=not_experimental
complement(14283)
/gene="GCTAI"
complement(14305..14310)
/gene="GCTAI"
14897..14934
/note="19 copies 2 mer aa 92% conserved"
complement(15755..17099)
/gene="GCTAI"
/note="match: STS: Em:G05861"
complement(16558..16912)
/gene="GCTAI"
/note="match: STS: Em:G59840"
complement(join(17100..17232,17420..17524,19637..19759,
24065..24252,38835..39070,70404..70815))
/gene="GCTAI"
/note="match: proteins: Tr:O60512"
/codon_start=1
/evidence=not_experimental
/product="ba326f20.1 (glycoprotein, alpha-galactosyl
transferase 1)"
/protein_id="CAD13306.1"
/db_xref="GI:17384434"
/translation="MRLREPLLSGSAMPAGSLQACRLNVAICALHLYVLYYLNG
RDLRLPOLVGVSTPLQGSNSAALGSGSELRGTGARGASDOPRGSSSP
VDSGPGPANTLSVPVHTTALSLPACPESEPLTGLPMLEFNMPLDELVAKNON
VKMGGRVAPRDQVSPHKVATITPERNOEHLKYIYHPLVLRQOCDGIVYINAG
DTIFNRKALLNVEFOELAKDYDTTCFYSVDLILPMNDHAYKCFSCPRITSANMF
GFSLPYVDFGVSAUSLKQDFITTCGPNMYWCGEDDIFNRLVFRGSISSPNDV
VGRCRMIRHSRDKNKPENPDRFRIAHTKETMLSDGLNLTGYLDVORYPLVYQITIV
DIGTPS"
repeat_region
17946..18109
/note="12 repeat: matches 2241..2411 of consensus"
repeat_region
19126..19243
/note="LIMC5 repeat: matches 7805..7927 of consensus"
repeat_region
19877..20157
/note="AlusC repeat: matches 11..299 of consensus"
repeat_region
20168..20468
/note="AlusX repeat: matches 1..301 of consensus"
repeat_region
22027..22341
/note="AluJo repeat: matches 4..309 of consensus"
repeat_region
22507..22807
/note="AlusX repeat: matches 1..301 of consensus"
repeat_region
23081..23376
/note="AluJo repeat: matches 1..297 of consensus"
repeat_region
23549..23852
/note="Aluub repeat: matches 1..305 of consensus"
repeat_region
24409..24625
/note="LINE repeat: matches 5615..5826 of consensus"

Query Match 65.2% Score 117.4; DB: 9; Length 163396;
Best Local Similarity 91.9% Pred. No. 5.3e-27;
Matches 124; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 7 CTCCTCTTCAGAAATGCCATCTGAGACACATGATAGCTCCACCTGTTCCAG 66
||||| |
Db 150244 CTCCTCTCTTCACACGCCCATCTGTGACACATGTAAGTCTCAACCTGTCCAG 150303
||||| |

QY 67 ATGTCACACTGCTGTGCGGACACTGATGGCTCATATATAGATGAATGCCACTTGC 126
||||| |
Db 150304 ATGTCACACTGCTGTGCGGACACTGATGGCTCATATATAGATGAATGCCACTTGC 150363
||||| |

QY 127 TTGGCCCGATMAAA 141
||||| |
Db 150364 TTGGCCCGATGTTAA 150378
||||| |

RESULT 14
AC027812/c
LOCUS
DEFINITION
Homo sapiens chromosome 9 clone RP11-366F19 map 9, WORKING DRAFT
SEQUENCE, 29 unordered pieces.
AC027812
AC027812.2 GI:7677927
HTG: HTGS PHASE1: HTGS_DRAFT.
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 176584)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 9, clone RP11-366F19
Unpublished
2 (bases 1 to 176584)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S.,
Dodg,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrum,T., Meneses,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tsfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 176584)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S.,
Dodg,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrum,T., Meneses,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tsfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 2, 2000 this sequence version replaced gi:17382654.
All repeats were identified using RepeatMasker:

OY 127 TTGGCCCGGATATAA 141
DB 136140 TTGGCCCGGATATAA 136126

RESULT 15
AC034144
LOCUS AC034144
DEFINITION Homo sapiens chromosome 9 clone RP11-126M6, WORKING DRAFT SEQUENCE,
5 unordered pieces.
ACCESSION AC034144
VERSION AC034144.5 GI:9929777
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 183099)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 183099)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 26, 2000 this sequence version replaced gi:9795673.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0126M06
----- Summary Statistics -----
Sequencing vector: Plasmid; 0%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 179431 bases at least Q40
Consensus quality: 180878 bases at least Q30
Consensus quality: 181543 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 182699; sum-of-ctnigs
Quality coverage: 6.27 in Q20 bases; agarose-fp
Quality coverage: 6.20 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 24313: contig of 24313 bp in length
* 24314 24413: gap of unknown length
* 24414 24520: contig of 30107 bp in length
* 54521 54620: gap of unknown length
* 54621 93200: contig of 38580 bp in length
* 93201 93300: gap of unknown length
* 93301 132512: contig of 39212 bp in length
* 132513 132612: gap of unknown length
* 132613 183099: contig of 50487 bp in length.
Location/Qualifiers
1. 183099
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"

misc_feature /clone="RP11-126M6"
1. 24313
misc_feature /note="assembly_name:Contig16"
24414. 54520
misc_feature /note="assembly_name:Contig17"
54621. 93200
misc_feature /note="assembly_name:Contig18"
93301. 132512
misc_feature /note="assembly_name:Contig19"
132613. 183099
BASE COUNT 52286 a 40806 c 40143 g 49464 t 400 others
ORIGIN

Query Match 65.2%; Score 117.4; DB 2; Length 183099;
Best Local Similarity 91.9%; Pred. No. 5.3e-27;
Matches 124; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 7 CTCCCTTCTCAGATGCCCATCTGTGACACATGTAGAGTCTCAACCTGTTCCAG 66
DB 64694 CTCCCTTCTCAGATGCCCATCTGTGACACATGTAGAGTCTCAACCTGTTCCAG 64753
OY 67 ATGTCCACCTGCTGTGGGACGTGATGGCTCACAATATGCAATGAATGCCAGCTTGC 126
DB 64754 ATGTCCACCTGCTGTGGGACGTGATGGCTCACAATATGCAATGAATGCCAGCTTGC 64813
OY 127 TTGGCCCGGATATAA 141
DB 64814 TTGGCCCGGATATAA 64828

Search completed: September 23, 2003, 12:11:12
Job time : 787.841 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:52:25 ; Search time 124.256 Seconds
(without alignments)
3910.464 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180
Sequence: 1 ggaagctcccttcctcaag.....tcattgaagatgccaatgc 180

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	100.0	323	24	ABV86895
2	180	100.0	323	24	ABV89019
3	180	100.0	389	19	AAV38073
4	180	100.0	394	24	ABV97573
5	180	100.0	398	18	AAV64673
6	180	100.0	402	20	AAZ77490
7	180	100.0	433	21	AAC98176
8	180	100.0	433	22	AAH34805

9	151.8	84.3	468	24	ABO60075
10	137.4	76.3	283	24	ABO60577
11	44.6	24.8	168	9	AAH80453
12	41.4	23.0	168	9	AAH80452
13	41.4	23.0	182	9	AAH80030
14	41.4	23.0	182	9	AAH80216
15	41.4	23.0	482	13	AAO24083
16	39.8	22.1	182	9	AAH82217
17	39.8	22.1	182	9	AAH82232
18	39.8	22.1	314	12	AAQ12109
19	39.8	22.1	415	17	AAH15982
20	38.8	21.6	182	9	AAH82227
21	38.8	21.6	182	9	AAH82228
22	38.8	21.6	182	9	AAH82233
23	38.2	21.2	171	11	AAQ01494
24	38.2	21.2	182	9	AAH82221
25	38.2	21.2	368	24	ABK6355
26	38.2	21.2	368	24	ABN97352
27	38.2	21.2	368	24	ABL64781
28	38.2	21.2	368	24	ABL65420
29	37.4	20.8	182	9	AAH82224
30	37.4	20.8	182	9	AAH82225
31	37.4	20.8	182	9	AAH82226
32	37.4	20.8	182	9	AAH82223
33	37.2	20.7	182	9	AAH82229
34	37.2	20.7	182	9	AAH82230
35	36.6	20.3	171	11	AAQ01577
36	36.6	20.3	179	9	AAH81535
37	36.6	20.3	182	9	AAH82222
38	36.6	20.3	200	10	AAH91164
39	36.6	20.3	200	12	AAQ14261
40	36.6	20.3	234	10	AAH91165
41	36.6	20.3	234	12	AAQ14262
42	36.6	20.3	238	18	AAH84523
43	36.6	20.3	302	15	AAQ78611
44	36.6	20.3	341	9	AAH80460
45	36.6	20.3	362	24	ABO60575

ALIGNMENTS

RESULT 1
ID ABV86895 standard; cDNA: 323 BP.
AC ABV86895;
XX
XX 13-DEC-2002 (first entry)
DE Human colon cancer related cDNA seq ID NO 206.
XX
XX Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KW ss.
XX Homo sapiens.
OS
XX
XX WO200258334-A2.
XX
XX PD 01-AUG-2002.
XX
XX 19-NOV-2001; 2001WO-US43704.
XX
XX 20-NOV-2000; 2000US-252222P.
PR 06-FEB-2001; 2001US-267011P.
PR 28-MAR-2001; 2001US-279670P.
PR 10-JUL-2001; 2001US-304037P.
XX
XX (CORI-) CORIXA CORP.
XX Stolk JA, Xu J, Chenaault RA, Meagher MJ, Secrist H, King GE;
XX WPI; 2002-608400/65.
XX
XX

XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 PS Claim 1: SEQ ID NO 206; 266pp + Sequence Listing: English.
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (I) can be used in gene therapy and
 CC (I) and (II) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 323 BP; 66 A; 81 C; 93 G; 83 T; 0 other;

Query Match 100.0%; Score 180; DB 24; Length 323;
 Best Local Similarity 100.0%; Pred. No. 1.1e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCACCTGT 60
 DB 217 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCACCTGT 158
 QY 61 TCCCAATATCCCAACTGGTCTGGCGGCTGATGGGCTCACATATCGAATGAATCCAG 120
 DB 157 TCCCAATATCCCAACTGGTCTGGCGGCTGATGGGCTCACATATCGAATGAATCCAG 98
 QY 121 CTCTGCTTGGCCCGGATTAACCAACAGAGCATCATGATCAATGAAGTGC 180
 DB 97 CTCTGCTTGGCCCGGATTAACCAACAGAGCATCATGATCAATGAAGTGC 38

RESULT 2
 AAV89019
 ID AAV89019 standard; cDNA; 323 BP.

AC AAV89019;

DT 13-DEC-2002 (first entry)

XX Human colon cancer related cDNA SEQ ID NO 2330.

KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;

KW ss.

OS Homo sapiens.

PN WO200258534-A2.

PD 01-AUG-2002.

PF 19-NOV-2001; 2001WO-US43704.

PR 20-NOV-2000; 2000US-252222P.

PR 06-FEB-2001; 2001US-267011P.

PR 28-MAR-2001; 2001US-279670P.

PR 10-JUL-2001; 2001US-304037P.

PA (CORI-) CORIXA CORP.

PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;

DR WPI; 2002-608400/65.

PT New isolated tumor colon polynucleotide and polypeptide, useful for the

PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 PS Claim 1: SEQ ID NO 2330; 266pp + Sequence Listing: English.

CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (I) can be used in gene therapy and
 CC (I) and (II) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 323 BP; 83 A; 93 C; 81 G; 66 T; 0 other;

Query Match 100.0%; Score 180; DB 24; Length 323;
 Best Local Similarity 100.0%; Pred. No. 1.1e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCACCTGT 60
 DB 107 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCACCTGT 166
 QY 61 TCCCAATATCCCAACTGGTCTGGCGGCTGATGGGCTCACATATCGAATGAATCCAG 120
 DB 167 TCCCAATATCCCAACTGGTCTGGCGGCTGATGGGCTCACATATCGAATGAATCCAG 226
 QY 121 CTCTGCTTGGCCCGGATTAACCAACAGAGCATCATGATCAATGAAGTGC 180
 DB 227 CTCTGCTTGGCCCGGATTAACCAACAGAGCATCATGATCAATGAAGTGC 286

RESULT 3
 AAV38073
 ID AAV38073 standard; cDNA; 389 BP.

AC AAV38073;

DT 14-SEP-1998 (first entry)

XX Human tumour-associated Kazal inhibitor encoding cDNA.

KW Human; tumour-associated Kazal inhibitor; TAKI; inflammation; cancer;

KW ulcerative colitis; Crohn's disease; inflammatory cytokine; elastase;

KW serine protease; glycoprotein hormone; trypsin; chymotrypsin; pepsin; ss.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 34..294

FT /tag-a /Product="tumour-associated Kazal inhibitor"

PN WO9820132-A1.

PD 14-MAY-1998.

PF 05-NOV-1997; 97WO-US20204.

PR 06-NOV-1996; 96US-0744670.

PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Goli SK, Murry LE;

DR WPI; 1998-286948/25.

DR P-PDB; AAW62074.

PT	New isolated polynucleotide and pancreatic tumor polypeptides, useful
PT	for diagnosing, preventing and/or treating cancer, particularly
PT	pancreatic cancer -
XX	
PS	Claim 1; SEQ ID NO 2981; 300pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated polynucleotide (I) comprising: (a)
CC	any of a group of over 4000 nucleotide sequences (ABV94678-ABV99145);
CC	(b) complements of (a); (c) sequences consisting of at least 20
CC	contiguous residues of (a); (d) sequences that hybridize to (a), under
CC	moderately stringent conditions; (e) sequences having at least 75% or 90%
CC	identity to (a); or (f) degenerate variants of (a). Polypeptides
CC	encoded by (I) and oligonucleotide can be used to
CC	detect cancer in a patient and compositions comprising polypeptides,
CC	polynucleotides, antibodies, fusion proteins, T cell populations and
CC	antigen presenting cells expressing the polypeptide are useful in
CC	treating pancreatic cancer and stimulating an immune response. The
CC	polynucleotides can be used as probes or primers for nucleic acid
CC	hybridisation. In the design and preparation of ribozyme molecules for
CC	inhibiting expression of the tumour polypeptides and proteins in the
CC	tumour cells, in vaccines and for gene therapy.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pcr_sequences.
XX	
SQ	Sequence 394 BP; 115 A; 102 C; 100 G; 73 T; 4 other:
Query Match	100.0%; Score 180; DB 24; Length 394;
Best Local Similarity	100.0%; Pred. No. 1.2e-51;
Matches 180; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	1 GGAAAGCTCCCTTCTCAAGAATGCCATCTGTGAACACANTGGTAGGTCCTCAACTGT 60 DB 77 GGAAAGCTCCCTTTCTCAAGAATGCCCATCTGTGAACACATGGTAGGTCCTCAACTGT 136
OY	61 TCCAGATGTCCAACCTCGTGCGGCACCTGATGGGCTCACATAAGCATAATGCCAG 120 DB 137 TCCAGATGTCCAACCTCGTGCGGCACCTGATGGGCTCACATAAGCATAATGCCAG 196
OY	121 CTCGCTGTGGCGCCGATATAAACCACAAGACATCCAGTCACTGAAGATGGCAAATGC 180 DB 197 CTCGCTGTGGCGCCGATATAAACCACAAGACATCCAGTCACTGAAGATGGCAAATGC 256
RESULT 5	
AAT64673	
ID	AAT64673 standard; cDNA to mRNA; 398 BP.
XX	
AC	AAT64673;
XX	
DT	22-Dec-1997 (first entry)
XX	
DE	Human PEC-60-type protein encoding cDNA.
XX	
KW	Human; PEC-60-type protein; digestive disorder; immune disorder; nervous disorder; gastrointestinal tract hormone; gastric tissue; antigen; antibody; ds.
KW	
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FH	CDS 43..303
FT	/*tag= a
FT	/product= PEC-60-type-protein
FT	sig_peptide 43..120
FT	/*tag= b
FT	mat_peptide 121..300
FT	/*tag= c
XX	
RN	W09715596-A1.
XX	
DD	01-MAY-1997.

[illegible]

PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-587534/55.
 DR P-PSDB: AAB53419.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 PS Claim 1: Page 610; 2104pp; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53334 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnerary, nephrotoxic, anti-infective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 CC
 XX Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;
 SQ
 Query Match 100.0%; Score 180; DB 21; Length 433;
 Best Local Similarity 100.0%; Pred. No. 1.2e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGACTCTCCAACTGT 60
 DB 137 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGACTCTCCAACTGT 196
 QY 61 TCCGAGATGCCAAGCTGTGCGGCACTGATGGGCTGCATATACGAATGAATGCCAG 120
 DB 137 TCCGAGATGCCAAGCTGTGCGGCACTGATGGGCTGCATATACGAATGAATGCCAG 256
 QY 121 CTCTGCTTGGCCGGGATTAACCAACAGGACATCCAGATCATCAAAAGATGGCAATGC 180
 DB 257 CTCTGCTTGGCCGGGATTAACCAACAGGACATCCAGATCATCAAAAGATGGCAATGC 316

RESULT 8
 AAH34805
 ID AAH34805 standard; cDNA; 433 BP.
 XX
 AC AAH34805;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1887.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 XX
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HDMA-) HUMAN GENOME SCI INC.
 XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI: 2001-235957/24.
 DR P-PSDB: AAG75400.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1: Page 3403; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAG77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 XX Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;
 SQ
 Query Match 100.0%; Score 180; DB 22; Length 433;
 Best Local Similarity 100.0%; Pred. No. 1.1e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGACTCTCCAACTGT 60
 DB 137 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGACTCTCCAACTGT 196
 QY 61 TCCGAGATGCCAAGCTGTGCGGCACTGATGGGCTGCATATACGAATGAATGCCAG 120
 DB 137 TCCGAGATGCCAAGCTGTGCGGCACTGATGGGCTGCATATACGAATGAATGCCAG 256
 QY 121 CTCTGCTTGGCCGGGATTAACCAACAGGACATCCAGATCATCAAAAGATGGCAATGC 180
 DB 257 CTCTGCTTGGCCGGGATTAACCAACAGGACATCCAGATCATCAAAAGATGGCAATGC 316

RESULT 9
 ABO60075
 ID ABO60075 standard; cDNA; 468 BP.
 XX
 AC ABO60075;
 XX
 DT 02-AUG-2002 (first entry)
 XX
 DE Human colon cancer related nucleotide sequence SEQ ID NO:3770.
 XX
 KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
 KW genetic analysis; diagnostic; antisense therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200229086-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 02-OCT-2001; 2001WO-US30732.
 XX
 PR 02-OCT-2000; 2000US-237271P.
 XX
 PA (FARB) BAYER CORP.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 11:18:55; Search time 33.3218 Seconds
(without alignments)
2384.294 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180
Sequence: 1 ggaagctcccttctcaag.....tcataaagatgcaaatgc 180

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCBUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	100.0	180	US-09-065-019-1	Sequence 1, Appli
2	180	100.0	388	US-08-744-670-2	Sequence 2, Appli
3	180	100.0	388	US-09-149-933-2	Sequence 2, Appli
4	180	100.0	398	US-09-065-019-2	Sequence 2, Appli
5	39.8	22.1	458	US-08-880-829-21	Sequence 21, Appli
6	36.6	20.3	302	US-08-235-515A-26	Sequence 26, Appli
7	35.4	19.7	432	US-09-016-434-1250	Sequence 1250, Ap
8	35.4	19.7	420	US-09-221-298-56	Sequence 56, Appli
9	34.4	19.1	285	US-09-016-434-361	Sequence 361, App
10	32.2	17.9	265	US-09-221-298-55	Sequence 55, Appli
11	32	17.8	309	US-08-839-709-2	Sequence 2, Appli
12	31.4	17.8	309	US-09-204-859-2	Sequence 2, Appli
13	31.4	17.4	1129	US-09-383-586-10	Sequence 10, Appli
14	31.4	17.4	1206	US-09-312-283C-357	Sequence 357, App
15	31	17.2	126	US-08-586-676E-4	Sequence 4, Appli
16	31	17.2	129	US-08-586-676E-5	Sequence 5, Appli
17	31	17.2	138	US-08-586-676E-6	Sequence 6, Appli
18	30.6	17.0	3112	US-09-318-448-9	Sequence 9, Appli
19	29.4	16.3	180	US-09-065-019-1	Sequence 1, Appli
20	29.4	16.3	388	US-08-744-670-2	Sequence 2, Appli
21	29.4	16.3	388	US-09-149-933-2	Sequence 2, Appli
22	29.4	16.3	398	US-09-065-019-2	Sequence 2, Appli
23	29	16.1	276	US-08-712-948-3	Sequence 3, Appli
24	28.8	16.0	1230025	US-09-198-452A-1	Sequence 1, Appli
25	28.4	15.8	463	US-09-439-313-362	Sequence 362, App
26	28.4	15.8	463	US-09-352-616A-362	Sequence 362, App
27	28.2	15.7	466	US-09-141-027-5	Sequence 5, Appli

28	28.2	15.7	466	4	US-09-617-804-5	Sequence 5, Appli
29	28.2	15.7	771	2	US-08-972-008-3	Sequence 3, Appli
30	28.2	15.7	771	4	US-09-267-409-3	Sequence 3, Appli
31	28.2	15.7	2495	4	US-09-141-027-1	Sequence 1, Appli
32	28.2	15.7	2495	4	US-09-617-804-1	Sequence 1, Appli
33	28.2	15.7	2525	2	US-08-972-008-1	Sequence 1, Appli
34	28.2	15.7	2525	4	US-09-267-409-1	Sequence 1, Appli
35	27.4	15.2	2226	3	US-08-951-260A-1	Sequence 1, Appli
36	27.4	15.2	2226	4	US-09-430-626A-1	Sequence 1, Appli
37	27.4	15.2	2949	4	US-09-412-554A-3	Sequence 3, Appli
38	27.2	15.1	1529	3	US-08-821-278A-1	Sequence 1, Appli
39	27.2	15.1	2818	4	US-09-221-017B-506	Sequence 506, App
40	27	15.0	1988	2	US-08-712-948-8	Sequence 8, Appli
41	26.8	14.9	17000	4	US-09-679-299A-18	Sequence 18, Appli
42	26.6	14.8	2511	4	US-09-252-991A-9494	Sequence 9494, Ap
43	26.6	14.8	2583	4	US-09-252-991A-9541	Sequence 9541, Ap
44	26.6	14.8	4531	4	US-09-620-312D-893	Sequence 893, App
45	26.2	14.6	254	4	US-09-016-434-313	Sequence 313, App

ALIGNMENTS

RESULT 1
US-09-065-019-1
Sequence 1, Application US/09065019
Patent No. 6280968
GENERAL INFORMATION:
APPLICANT: Kato, Selsni
APPLICANT: Yamaguchi, Tomoko
APPLICANT: Sekine, Shingo
APPLICANT: Kamata, Kouju
TITLE OF INVENTION: HUMAN PEC-60-LIKE PROTEIN AND DNA ENCODING THIS PROTEIN
FILE REFERENCE: 6700PCT-US
CURRENT APPLICATION NUMBER: US/09/065, 019
CURRENT FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 180
TYPE: DNA
ORGANISM: Homo sapiens
US-09-065-019-1

Query Match 100.0%; Score 180; DB 3; Length 180;
Best Local Similarity 100.0%; Pred. No. 3.2e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGTAACACATGTAGATCTCCCAACTGT 60
DB 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGTAACACATGTAGATCTCCCAACTGT 60
TCCAGATGTCACACCTGTGCGGACCTGATGCGGTACATATACGAATGCGCAG 120
TCCAGATGTCACACCTGTGCGGACCTGATGCGGTACATATACGAATGCGCAG 120
61 TCCAGATGTCACACCTGTGCGGACCTGATGCGGTACATATACGAATGCGCAG 120
DB 61 TCCAGATGTCACACCTGTGCGGACCTGATGCGGTACATATACGAATGCGCAG 120
121 CTCTGCTTGCCCGGATATAAACCACAGACATCCAGATCAGTAAAGATGGCAATGC 180
DB 121 CTCTGCTTGCCCGGATATAAACCACAGACATCCAGATCAGTAAAGATGGCAATGC 180

RESULT 2
US-08-744-670-2
Sequence 2, Application US/08744670
Patent No. 5858710
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
APPLICANT: Murty, Lynn E.
TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

6, 280, 918

15/9/03

STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/744,670
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0155 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINTTUT01
CLONE: 1539065
US-08-744-670-2

Query Match 100.0%; Score 180; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 4.5e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
DB 111 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 170
QY 61 TCCAGATGTCACACCTGCTCGGGCAGCTGATGGGCTCAATATACGAATGAATGCCAG 120
DB 171 TCCAGATGTCACACCTGCTCGGGCAGCTGATGGGCTCAATATACGAATGAATGCCAG 230
QY 121 CTCTGCTTGGCCGGATAAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 180
DB 231 CTCTGCTTGGCCGGATAAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 290

RESULT 3

US-09-149-933-2
Sequence 2, Application US/09149933
Patent No. 5938699
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
APPLICANT: Murty, Lynn E.
TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/149,933
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0155 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINTTUT01
CLONE: 1539065
US-09-149-933-2

Query Match 100.0%; Score 180; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 4.5e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
DB 111 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 170
QY 61 TCCAGATGTCACACCTGCTCGGGCAGCTGATGGGCTCAATATACGAATGAATGCCAG 120
DB 171 TCCAGATGTCACACCTGCTCGGGCAGCTGATGGGCTCAATATACGAATGAATGCCAG 230
QY 121 CTCTGCTTGGCCGGATAAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 180
DB 231 CTCTGCTTGGCCGGATAAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 290

RESULT 4

US-09-065-019-2
Sequence 2, Application US/09065019
Patent No. 6280968
GENERAL INFORMATION:
APPLICANT: Kato, Seishi
APPLICANT: Yamaguchi, Tomoko
APPLICANT: Sekine, Shingo
APPLICANT: Kamata, Kouju
TITLE OF INVENTION: HUMAN PEC-60-LIKE PROTEIN AND DNA ENCODING THIS PROTEIN
FILE REFERENCE: 6700PCT-US
CURRENT APPLICATION NUMBER: US/09/065,019
CURRENT FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 398
TYPE: DNA
ORGANISM: Homo sapiens
US-09-065-019-2

Query Match 100.0%; Score 180; DB 3; Length 398;
Best Local Similarity 100.0%; Pred. No. 4.5e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
DB 121 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 180
QY 61 TCCAGATGTCACACCTGCTCGGGCAGCTGATGGGCTCAATATACGAATGAATGCCAG 120
DB 181 TCCAGATGTCACACCTGCTCGGGCAGCTGATGGGCTCAATATACGAATGAATGCCAG 240

OY 121 CTCGCTGGCCGGATTAACCAACAGACATCCAGATCATGAAGAATGCAATGC 180
Db 241 CTCGCTGGCCGGATTAACCAACAGACATCCAGATCATGAAGAATGCAATGC 300

RESULT 5
US-08-880-829-21

Sequence 21, Application US/08880829
Patent No. 5923559
GENERAL INFORMATION:
APPLICANT: Collins, John
APPLICANT: Roettgen, Peter
TITLE OF INVENTION: A Collection of Phagemids, A
TITLE OF INVENTION: Collection of Escherichia coli
TITLE OF INVENTION: Cells Carrying The Phagemids, A
TITLE OF INVENTION: Collection of Phagemid Particles
TITLE OF INVENTION: Produced From Said Collection
TITLE OF INVENTION: And Phagemid Particles
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Joseph T. Eisele
ADDRESSEE: Kane, Dalsimer, Sullivan, Kurucz,
ADDRESSEE: Levy, Eisele and Richard
STREET: 711 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017-4059
COMPUTER READABLE FORM:
MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-XT COMPATIBLE
OPERATING SYSTEM: DOS 3.3:
SOFTWARE: WORDPERFECT 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880, 829
FILING DATE: 23-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/458, 668
FILING DATE: 06/02/95
APPLICATION NUMBER: German EP 94 108 689.4
FILING DATE: 06/07/94
ATTORNEY/AGENT INFORMATION:
NAME: EISELE, JOSEPH T.
REGISTRATION NUMBER: 25,331
REFERENCE/DOCKET NUMBER: 2727-77
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 687-6000
TELEFAX: (212) 682-3485
TELEX: (212) 426767
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 base pairs
TYPE: nucleic acid
STRANDEDNESS: single strand
TOPOLOGY: linear
MOLECULE TYPE:
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
US-08-880-829-21

Query Match 22.1%; Score 39.8; DB 2; Length 458;
Best Local Similarity 57.7%; Pred. No. 6.2e-05;
Matches 71; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 58 TGTTCACGATGTCACCTGTCGCGGCGACGTGAGGCTCACATATACGAATGAATGC 117
Db 220 TGCACCTAAGATCTACGACCGCGTTTGGCGTACCGACGCAACACTTACCCGACGAATGC 279
OY 118 CAGCTCTGCTGGCCCGGATTAACCAACAGACATCCAGATCATGAAGAATGCAAA 177
Db 280 GTTCTGCTGCTGAACCAACCGTAACGTCAGACTCTATCTGATCCAGAAATCTGCTCG 339
OY 178 TGC 180
Db 340 TGC 342

RESULT 6

US-08-235-515A-26
Sequence 26, Application US/08235515A
Patent No. 5840518
GENERAL INFORMATION:
APPLICANT: Morishita, Hideaki
APPLICANT: Kanemori, Toshinori
TITLE OF INVENTION: DNA FRAGMENT, VECTOR CONTAINING THE DNA
TITLE OF INVENTION: FRAGMENT, TRANSFORMANT TRANSFORMED WITH THE VECTOR AND
TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEIN USING THE VECTOR
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/235, 515A
FILING DATE: 29-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1110-135P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8050
TELEFAX: 703-205-8000
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: -
LOCATION: 1..302 /label= polynucleotide
OTHER INFORMATION: /note= "j-xi-y-z(pstl), insert in plasmid pm474,
OTHER INFORMATION: Figure 5"
FEATURE:
NAME/KEY: slg-peptide
LOCATION: 27..89
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 90..293
FEATURE:


```

; FILE REFERENCE: CE001470
; CURRENT APPLICATION NUMBER: US/60/485,450

```


score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

9394.424 Million cell updates/sec

Sequence: 1 gcagccccagccagctcag.....taaaagatccagcccaactg 398

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:	5777422
--	---------

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1:	GenEmbl:	1:	gb.ba.*
2:	gb.htg.*	2:	gb.htg.*
3:	gb.lin.*	3:	gb.lin.*
4:	gb.om.*	4:	gb.om.*
5:	gb.ov.*	5:	gb.ov.*
6:	gb.pat.*	6:	gb.pat.*
7:	gb.ph.*	7:	gb.ph.*
8:	gb.pl.*	8:	gb.pl.*
9:	gb.pr.*	9:	gb.pr.*
10:	gb.ro.*	10:	gb.ro.*
11:	gb.sts.*	11:	gb.sts.*
12:	gb.sy.*	12:	gb.sy.*
13:	gb.un.*	13:	gb.un.*
14:	gb.vi.*	14:	gb.vi.*
15:	em.ba.*	15:	em.ba.*
16:	em.fun.*	16:	em.fun.*
17:	em.hum.*	17:	em.hum.*
18:	em.in.*	18:	em.in.*
19:	em.mu.*	19:	em.mu.*
20:	em.om.*	20:	em.om.*
21:	em.or.*	21:	em.or.*
22:	em.ov.*	22:	em.ov.*
23:	em.pat.*	23:	em.pat.*
24:	em.ph.*	24:	em.ph.*
25:	em.pl.*	25:	em.pl.*
26:	em.ro.*	26:	em.ro.*
27:	em.sts.*	27:	em.sts.*
28:	em.un.*	28:	em.un.*
29:	em.vi.*	29:	em.vi.*
30:	em.hsg.hum.*	30:	em.hsg.hum.*
31:	em.hsg.in.v.*	31:	em.hsg.in.v.*
32:	em.hsg.other.*	32:	em.hsg.other.*
33:	em.hsg.mus.*	33:	em.hsg.mus.*
34:	em.hsg.pln.*	34:	em.hsg.pln.*
35:	em.hsg.rod.*	35:	em.hsg.rod.*
36:	em.hsg.wam.*	36:	em.hsg.wam.*
37:	em.hsg.vrt.*	37:	em.hsg.vrt.*
38:	em.hsg.*	38:	em.hsg.*
39:	em.hsgo.hum.*	39:	em.hsgo.hum.*
40:	em.hsgo.mus.*	40:	em.hsgo.mus.*
41:	em.hsgo.other.*	41:	em.hsgo.other.*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	398	100.0	398	6	AR166138	AR166138 Sequence
2	398	100.0	398	6	E13091	E13091 Human cDNA
3	398	100.0	402	6	AX014851	AX014851 Sequence
4	388	97.5	388	6	AR028526	AR028526 Sequence
5	388	97.5	388	6	AR075817	AR075817 Sequence
6	388	97.5	388	6	BD007521	BD007521 Tumor-ass
7	367	92.2	386	9	AF048700	AF048700 Homo sapi
8	248.8	62.5	364	4	S46866	S46866 PRC-60-gast
9	248.8	62.5	364	4	SSEPC60	S46866 PRC-60-gast
10	211.2	53.1	515	10	MMMPG60	Y11505 M.musculus
11	180	45.2	180	6	AR166137	AR166137 Sequence
12	180	45.2	180	6	E13090	E13090 Human an cd
13	145.8	36.6	163396	9	AL161445	AL161445 Human DNA
14	145.8	36.6	176384	2	AC027812	AC027812 Homo sapi
15	145.8	36.6	183039	2	AC034144	AC034144 Homo sapi
16	117.4	29.5	188540	2	AC011108	AC011108 Homo sapi
17	81.2	20.4	192093	2	AC121305	AC121305 Rattus no
18	81.2	20.4	230493	10	AL837521	AL837521 Mouse DNA
19	81.2	20.4	242175	2	AC096463	AC096463 Rattus no
20	45.6	11.5	573	6	BD027885	BD027885 Sequence
21	45.6	11.5	621	6	BD027884	BD027884 Sequence
22	44.6	11.2	182	6	A11350	A11350 Artificial
23	44.6	11.2	182	6	A11352	A11352 Artificial
24	44.6	11.2	182	6	A24905	A24905 PSTI master
25	44.6	11.2	182	12	SYNPSTIAA	M21832 Synthetic h
26	44.6	11.2	186	6	A24907	A24907 PSTI master
27	42.2	10.6	368	6	AX332609	AX332609 Sequence
28	42.2	10.6	368	6	AX333248	AX333248 Sequence
29	42.2	10.6	368	6	AX411203	AX411203 Sequence
30	42.2	10.6	368	6	BD160949	BD160949 Preventiv
31	42.2	10.6	368	9	HSPST1	Y00705 Homo sapien
32	41.4	10.4	168	6	T08731	T08731 Sequence
33	41.4	10.4	482	6	E03479	E03479 cDNA encod1
34	41.4	10.4	5635	12	ASVPSKAN8	X84307 Artificial
35	41.2	10.4	369	10	MMPIP12	X06342 Mouse RNA
36	40.6	10.2	341	6	E01574	E01574 cDNA sequen
37	40.6	10.2	362	9	BC025790	BC025790 Homo sapi
38	40.6	10.2	432	6	AR270687	AR270687 Sequence
39	40.6	10.2	432	6	AR270687	AL11949 Human pancr
40	40.6	10.2	608	6	AX014216	AX014216 Sequence
41	40.2	10.2	2000	6	AX655393	AX655393 Sequence
42	40.2	10.1	2228	10	AF094609	AF094609 Rattus no
43	39.6	9.9	179	6	E01725	E01725 Synthetic D
44	39.6	9.9	179	6	E02456	E02456 DNA encodin
45	39.6	9.9	204	12	SYNPSTIB	M31103 Synthetic h

ALIGNMENTS

RESULT	1
LOCUS	AR166138
DEFINITION	Sequence 2 from patent US 6280968.
ACCESSION	AR166138
VERSION	AR166138.1 GI:16241313
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 398)
TITLE	Kato,S., Yamaguchi,T., Sekine,S. and Kamata,K.
JOURNAL	Human PEC-60-like protein and DNA encoding the same
FEATURES	Patent: US 6280968-A 2 28-AUG-2001;
	Location/Qualifiers

source 1. .398
 BASE COUNT 107 a 110 c 105 g 76 t
 ORIGIN

Query Match 100.0%; Score 398; DB 6; Length 398;
 Best Local Similarity 100.0%; Pred. No. 2.8e-99;
 Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGCCCGCAGCTCAGGCTACACTATCCAGATCAGCATGGCCGCTCCAGTGG 60
 DB 1 GCAGGCCCGCAGCTCAGGCTACACTATCCAGATCAGCATGGCCGCTCCAGTGG 60
 QY 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGGAAGTCCAGTGGCAGCA 120
 DB 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGGAAGTCCAGTGGCAGCA 120
 QY 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGAGTCTCAACCTGT 180
 DB 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGAGTCTCAACCTGT 180
 QY 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAG 240
 DB 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAG 240
 QY 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATATGAAGATGGCAATGC 300
 DB 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATATGAAGATGGCAATGC 300
 QY 301 TGAATCCACAGAGGACCTCTCAAGCAATGCCATGTCACTGAGAGAGTGGTGGCATGG 360
 DB 301 TGAATCCACAGAGGACCTCTCAAGCAATGCCATGTCACTGAGAGAGTGGTGGCATGG 360
 QY 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398
 DB 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398

RESULT 2
 LOCUS E13091 398 bp DNA linear PAT 27-APR-1998
 DEFINITION Human cDNA encoding a PEC-60-like protein.
 ACCESSION E13091
 VERSION E13091.1 GI:3251903
 KEYWORDS JP 1997124698-A/2.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 398)
 Kato,M., Yamaguchi,T., Sekine,S. and Kamata,M.
 HUMAN PEC-60-LIKE PROTEIN AND DNA CAPABLE OF CODING THE SAME
 Patent: JP 1997124698-A 2 13-MAY-1997;
 SAGAMI CHEM RES CENTER
 OS Homo sapiens (human)
 PN JP 1997124698-A/2
 PD 13-MAY-1997
 PE 27-OCT-1995 JP 1995280272
 PI KATO MASASHI, YAMAGUCHI TOMOKO, SEKINE SHINGO, PI KAMATA
 MITSUHIKA
 PC C07K14/575, C07H21/04, C12N15/09//A61K38/22, A61K38/22, A61K38/22,
 C12P21/02,
 PC (C12P21/02, C12R1:19);
 CC strandedness: Double;
 CC topology: Linear;
 FH Key Location/Qualifiers
 FT source 1. .398
 FT /organism='Homo sapiens'
 FT /tissue_type='stomach cancer' FT CDS
 FT 43. .303 /product='PEC-60-like protein precursor' FT
 FT sig_peptide 43. .120

FT mat_peptide 121. .300
 FT /product='PEC-60-like protein'.
 FEATURES Location/Qualifiers
 source 1. .398
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'

BASE COUNT 107 a 110 c 105 g 76 t
 ORIGIN

Query Match 100.0%; Score 398; DB 6; Length 398;
 Best Local Similarity 100.0%; Pred. No. 2.8e-99;
 Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGCCCGCAGCTCAGGCTACACTATCCAGATCAGCATGGCCGCTCCAGTGG 60
 DB 1 GCAGGCCCGCAGCTCAGGCTACACTATCCAGATCAGCATGGCCGCTCCAGTGG 60
 QY 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGGAAGTCCAGTGGCAGCA 120
 DB 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGGAAGTCCAGTGGCAGCA 120
 QY 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGAGTCTCAACCTGT 180
 DB 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGAGTCTCAACCTGT 180
 QY 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAG 240
 DB 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAG 240
 QY 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATATGAAGATGGCAATGC 300
 DB 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATATGAAGATGGCAATGC 300
 QY 301 TGAATCCACAGAGGACCTCTCAAGCAATGCCATGTCACTGAGAGAGTGGTGGCATGG 360
 DB 301 TGAATCCACAGAGGACCTCTCAAGCAATGCCATGTCACTGAGAGAGTGGTGGCATGG 360
 QY 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398
 DB 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398

RESULT 3
 LOCUS AX014851 402 bp DNA linear PAT 07-SEP-2000
 DEFINITION Sequence 41 from Patent WO9953040.
 ACCESSION AX014851
 VERSION AX014851.1 GI:10041118
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1
 Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
 Pillarsky,C.
 Human nucleic acid sequences from ovarian tumour tissue
 Patent: WO 9953040-A 41 21-OCT-1999;
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
 BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
 (DE); PILLARSKY CHRISTIAN (DE)
 Location/Qualifiers
 source 1. .402
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'

BASE COUNT 108 a 111 c 107 g 76 t
 ORIGIN

Query Match 100.0%; Score 398; DB 6; Length 402;
 Best Local Similarity 100.0%; Pred. No. 2.8e-99;
 Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match	97.5%;	Score 388;	DB 6;	Length 388;
Best Local Similarity	100.0%;	Prod. No. 1.6e-96;		
Matches 388;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	11	GCCAGCTCAGGCTACACTATCCAGAGTACAGCATGCGCCGCGGCATGGGTAATCGCCC	70	
Db	1	GCCAGCTCAGGCTACACTATCCAGAGTACAGCATGCGCGCTCGGCAGTGGGTAATCGCCC	60	
QY	71	TGGCCTTGCGCTGCCCTCCTTGTTGTGGACAGGGAAGTCCAGTGGCAGCAGAAAGCTCC	130	
Db	61	TGGCCTTGCGCTGCCCTCCTTGTTGTGGACAGGGAAGTCCAGTGGCAGCAGAAAGCTCC	120	
QY	131	CTTTCTCAAGATGCCCATCTGTGAACACATGATGATCTCCAAACCTGTGCCAGAGT	190	
Db	121	CTTTCTCAAGATGCCCATCTGTGAACACATGATGATCTCCAAACCTGTGCCAGAGT	180	
QY	191	CCAACCTGTGCTGGGGCACTGATGGGGCTACATATACGAATGAATGCAGCTTCGCTTGG	250	
Db	181	CCAACCTGTGCTGGGGCACTGATGGGGCTACATATACGAATGAATGCAGCTTCGCTTGG	240	
QY	251	CCCGGATAAAACCAACAGGACATCCAGATCATGAAAAGTGCCAAATGCTGATCCACA	310	

Query Match	Best Local Similarity	97.5%; Score 388; DB 6; Length 388;	100.0%; Pred. No. 1,6e-96;	Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	11	GCCAGCTCAGGCTACACTATCCACAGATCAGCATGCGCGTCCGCCAGTGGGTATCGCCC	70	
Db	1	GCCAGCTCAGGCTACACTATCCACAGATCAGCATGCGCGTCCGCCAGTGGGTATCGCCC	60	
QY	71	TGGCGCTTGGCTGGCCCTCTGTGTGTGTGACACAGGAAGTGGCCAGTGGCAGCAGGAAGCTCC	130	
Db	61	TGGCGCTTGGCTGGCCCTCTGTGTGTGTGACACAGGAAGTGGCCAGTGGCAGCAGGAAGCTCC	120	
QY	131	CTTTCTCAGAATGCCATGTGTGACACATGTTAGTCTCCAACTGTTCCAGATGT	190	
Db	121	CTTTCTCAGAATGCCATGTGTGACACATGTTAGTCTCCAACTGTTCCAGATGT	180	
QY	191	CCAACTGGTCTGGGGCAGTGTGGGTGCATCTACGATGAAATGATGGCAGCTGTGCTGG	250	
Db	181	CCAACTGGTCTGGGGCAGTGTGGGTGCATCTACGATGAAATGATGGCAGCTGTGCTGG	240	
QY	251	CCCGGATAAAAACCAACAGGACATCCAGATCATGAAAGATGGCAATGTCTGATCCACA	310	
Db	241	CCCGGATAAAAACCAACAGGACATCCAGATCATGAAAGATGGCAATGTCTGATCCACA	300	
QY	311	GGAGCACCCTCAAGCGCATGAAGTGTCTACCTGGAGAAACAGTGGTGGCATGGAGAGGATATG	370	
Db	301	GGAGCACCCTCAAGCGCATGAAGTGTCTACCTGGAGAAACAGTGGTGGCATGGAGAGGATATG	360	
QY	371	ACATGAATAAAGATCCAGCCCAACTG	398	
Db	361	ACATGAATAAAGATCCAGCCCAACTG	388	
RESULT 6	BD007521			
LOCUS	BD007521	388 bp	DNA	linear
DEFINITION	Tumor-associated KAZAL type inhibitor.			
ACCESSION	BD007521			
VERSION	BD007521.1	GI:18635894		
KEYWORDS	UP 2001503629-A/1.			
SOURCE	unidentified			
ORGANISM	unidentified			

REFERENCE 1 (bases 1 to 388)
AUTHORS Bandman, O., Goli, S.K. and Murry, L.E.
TITLE Tumor-associated KAZAL type inhibitor
JOURNAL Patent: JP 2001503629-A 1 21-MAR-2001;
INCYTE PHARMACEUTICALS INC
COMMENT OS Unidentified
PN JP 2001503629-A/1
PD 21-MAR-2001
PR 05-NOV-1997 JP 1998521770
PR 06-NOV-1996 US 08/7744670
PI OIGA BANDMAN, SURYA K GOLI, LYNN E MURRY
PC C12N15/09, A61K38/55, A61K39/395, A61P1/00, A61P35/00, C07K14/81,
PC C07K16/38,
PC C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12Q1/68, G01N33/574, PC
C12N15/00, C12N5/00,
PC A61K37/64
CC
FH
FT
FT
KEY Location/Qualifiers
Location/Qualifiers
1..388
/organism="unidentified".
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN

Query Match 97.5%; Score 388; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.6e-96;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCCAGCTCAGCTACACTATCCAGATCAGATGCGCCGTCGCGCAATGGGTAATCGCCC 70
DB 1 GCCAGCTCAGCTACACTATCCAGATCAGATGCGCCGTCGCGCAATGGGTAATCGCCC 60
QY 71 TGGCCTTGCTGCTCCCTCTCTTGTGTGAGACAGGGAAGTGCAGAGGCAAGAAAGCTCC 130
DB 61 TGGCCTTGCTGCTCCCTCTCTTGTGTGAGACAGGGAAGTGCAGAGGCAAGAAAGCTCC 120
QY 131 CTTTTCGAAGAATGCCCATATCTGTGAACACATGTAAGTGTCCCACTGTTCCCAATGT 190
DB 121 CTTTTCGAAGAATGCCCATATCTGTGAACACATGTAAGTGTCCCACTGTTCCCAATGT 180
QY 191 CCAACCTGCTGCTGGGAGCTGATGGGCTCATATATAGATGAATGCGAGCTGCTTGG 250
DB 181 CCAACCTGCTGCTGGGAGCTGATGGGCTCATATATAGATGAATGCGAGCTGCTTGG 240
QY 251 CCCGATAAACCAACACAGAGATCCAGATCATGAAGAATGGCAATGCTGATCCACA 310
DB 241 CCCGATAAACCAACACAGAGATCCAGATCATGAAGAATGGCAATGCTGATCCACA 300
QY 311 GGAGCACCCTCAACGATGAAGTGTCACTGGAGAACAGTGTGGGCTGAGAGAGATATG 370
DB 301 GGAGCACCCTCAACGATGAAGTGTCACTGGAGAACAGTGTGGGCTGAGAGAGATATG 360
QY 371 ACATGAATTAAGAATCCAGCCCACTG 398
DB 361 ACATGAATTAAGAATCCAGCCCACTG 388

RESULT 7
AF048700 386 bp mRNA linear PRI 06-MAR-1998
LOCUS Homo sapiens gastrointestinal peptide (P6C-60) mRNA, complete cds.
DEFINITION
ACCESSION AF048700
VERSION AF048700.1 GI:2935439
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 386)
AUTHORS Hu, G.
TITLE Human gastrointestinal peptide mRNA, homolog of swine PEC-60
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 386)
AUTHORS Hu, G.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1998) Shanghai Institute of Cell Biology, 320
Yue-Yang Road, Shanghai 200031, China
COMMENT Location/Qualifiers
1..386
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..386
/gene="PEC-60"
12..272
/gene="PEC-60"
/note="similar to Sus scrofa PEC-60"
/codon_start=1
/product="gastrointestinal peptide"
/protein_id="AAC05124.1"
/db_xref="GI:2935440"
/translation="MAVRQVYLAALALVYDREVPVAAKLPFSRMPICEHWESP
TCSQMSLVCGTGTGTYTNECQLARIKTKDQIMKDKC"
BASE COUNT 119 a 97 c 98 g 72 t
ORIGIN

Query Match 92.2%; Score 367; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.1e-90;
Matches 367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CCAGATCAGCATGCGCGTCCGAGTGGGTAATCGCCCTGCGCTTGGCTGCTCCTTGG 91
DB 1 CCAGATCAGCATGCGCGTCCGAGTGGGTAATCGCCCTGCGCTTGGCTGCTCCTTGG 60
QY 92 TTGTGACAGGGAAGTCCAGTGGCAGCAGAGAAAGCTCCTTTCTCAAGATGCCATCT 151
DB 61 TTGTGACAGGGAAGTCCAGTGGCAGCAGAGAAAGCTCCTTTCTCAAGATGCCATCT 120
QY 152 GTGAACCATGTGTAAGTCTCCAACTGTTTCCAGATGTCCAACCTGTGTGGGCACTG 211
DB 121 GTGAACCATGTGTAAGTCTCCAACTGTTTCCAGATGTCCAACCTGTGTGGGCACTG 180
QY 212 ATGGGCTCAGATATAGCAATGCAATGCGAGCTGCTGCGCGGATTAACCAACAGG 271
DB 181 ATGGGCTCAGATATAGCAATGCAATGCGAGCTGCTGCGCGGATTAACCAACAGG 240
QY 272 ACATCCAGATCATGAAGAATGGCAAAATGCTGATCCACAGAGCACCCTCAAGCCATGAAG 331
DB 241 ACATCCAGATCATGAAGAATGGCAAAATGCTGATCCACAGAGCACCCTCAAGCCATGAAG 300
QY 332 TGTACGTGGAGAAACAGTGTGGGCTATGGAGAGATATGACATGAATTAATAATCCAGC 391
DB 301 TGTACGTGGAGAAACAGTGTGGGCTATGGAGAGATATGACATGAATTAATAATCCAGC 360
QY 392 CCAACTG 398
DB 361 CCAACTG 367

RESULT 8
S46866 364 bp mRNA linear MAM 05-AUG-1999
LOCUS PEC-60-gastrointestinal peptide [swine, duodenum, mRNA, 364 nt].
DEFINITION
ACCESSION S46866
VERSION S46866.1 GI:257597
KEYWORDS
SOURCE Sus scrofa (pig)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

AUTHORS		TITLE		JOURNAL		MEDLINE		REMARK		FEATURES	
Meesis,M., Cintra,A., Solfrini,V., Enforis,P., Bortolotti,F., Morasutti,D.G., Ostenson,C.G., Efendić,S., Agerberth,B., Mutt,V. et.al.,		Molecular cloning of PEC-60 and expression of its mRNA and peptide in the gastrointestinal tract and immune system		J. Biol. Chem. 267 (28), 19829-19832 (1992)		93015834		1400298		GenBank staff at the National Library of Medicine created this entry [NCBI gisbg 115614] from the original journal article. This sequence comes from Fig. 1.	
		Location/Qualifiers									
		1. .364									
		/organism="Sus scrofa"									
		/mol_type="mRNA"									
		/db_xref="taxon:9823"									
		1. .364									
		/gene="PEC-60"									
		23. .283									
		/gene="PEC-60"									
		/note="gastrointestinal peptide; This sequence comes from Fig. 1; Author includes translated amino acids, 5' of putative initiation codon"									
		/codon_start=1									
		/product="PEC-60"									
		/protein_id="AAB23691.2"									
		/db_xref="GI:5705936"									
		/translation="MAVRLMWALALALITVDREVPVSAEKQVFSMPICEHMTESP									
		DSCRIVPVCSTDVTEYSESEKCLCLAIENKQDIQIVKDGSC"									
BASE COUNT		91 a 93 c 108 g 72 t									
ORIGIN											
Query Match		62.5%; Score 248.8; DB 4; Length 364;									
Best Local Similarity		82.5%; Pred. No. 6.2e-58;									
Matches 297; Conservative		0; Mismatches 62; Indels 1; Gaps 1;									
OY	25	CACATATCCCAAGATCAGATGCGCGTCCCGCAGTGGTAATCGCCCTTGCGTGC	84								
Db	5	CGCCAGGCCCAAGATCAGATGCGCGTCCCGTGGGTGGTGGCCCTGGCTGC	64								
OY	85	CTCCTTGTGTGGAGCAGGGAAGTGCAGTGGCAGCAGGAAGGTCCTTCTCAAGATG	144								
Db	65	CTCTTCATTGTGGACAGAGAGTGTCCAGTGTGGCAGGAAGAGGTTTCTCAAGATG	124								
OY	145	CCCATCTGTGAACATGTTAGTCTTCAACCTGTTCCCGATGTCCAACTGCTGTC	204								
Db	125	CCCATCTGTGAACATGAGGAGTGTCCCGACAGTGTCCCGATGTAGACAGCTGCT	184								
OY	205	GGCAGTGTGGCTCAGATATAGCAATGATGTCAGCTCTGCTGGGCCGGATAAAAC	264								
Db	185	GGCAGATGTGGCTCAGATATAGCAATGATGTCAGCTCTGCTGGCTGGATGAAAC	244								
OY	265	AAACAGGACATCCAGATCATGAAGATGGCAAAATGCTGATCCACAGAGCAGCTCAAC	324								
Db	245	AAACAGGACATCCAGATCATGAAGATGGCAAAATGCTGATCCACAGAGCAGCTTAAGC	304								
OY	325	CATGAAGTGTG-AGCTGAGAGAACAGTGGTGGGCTGAGAGATGATGATGAATAAA	383								
Db	305	CATGAAGTGTG-AGCTGAGAGAACAGTGGTGGGCTGAGAGATGATGATGAATAAA	364								
RESULT 9											
LOCUS		SSPEC60									
DEFINITION		S.scrofa PEC-60 mRNA.									
ACCESSION		X67109									
KEYWORDS		X67109.1 GI:2033									
SOURCE		Sus scrofa (p19)									
ORGANISM		Sus scrofa									
REFERENCE		Mammalia; Euthera; Cetartiodactyla; Suidae; Sus.									
		1 (bases 1 to 364)									

AUTHORS	Metzls M.
TITLE	Direct Submission
JOURNAL	Submitted (22-JUN-1992) M. Metzls, Dept. of Medical Chemistry, Lab. of Molecular Neurobiology, Karolinska Institute, Box 604 00, 10401 Stockholm, SWEDEN
FEATURES	Location/Qualifiers
SOURCE	1. 364 /organism="Sus scrofa" /mol_type="mRNA" /db_xref="taxon:9823" /clone_1lb="pig duodenum cDNA library in gt10 4x10 6" 23. 283
CDS	/codon_start=1 /evidence="experimental" /product="peptide PEC-60" /protein_id="CAA47482.1" /db_xref="GI:2034" /db_xref="SWISS-PROT:P37109" /translation="MAVRLMVALALALFLVIREVYSAEKQFSPMPICEHMTESP DCSRIDPCGSDVTYESDEKCLALRIENKDIQVDEBC" 23. 76 /note="peptide PEC-60" 101. 280 /product="peptide PEC-60"
BASE COUNT	91 a 93 c 108 g 72 t
ORIGIN	
Query Match	62.5%; Score 248.8; DB 4; Length 364;
Best Local Similarity	82.5%; Pred. No. 6.2e-58;
Matches	297; Conservative 0; Mismatches 62; Indels 1; Gaps 1;
OY	25 CACATATCCAGGATACAGCAGTGGCCGTCGCCAGTGGGTATATGCGCCCTGGAGCTGCC 84
Db	
OY	5 CGCAGCGCCAGATACAGCATGGCTGCTCCGCTGGGCGTCCGCCCTGGAGCTGCC 64
Db	
OY	85 CTCCTTGTGTGGACAGGGAGTGGCCACTGGCAGCAGAAAGCTCCCTTCTCAAGATG 144
Db	
OY	65 CTCCTCATTTGGGACAGAGAAGTCCACAGTGGCAGAAAAGAGTTTCTCAAGATG 124
OY	145 CCCATCTGTGAACACATGTGAGTCTCCCACTCTTCCCGAGTGCACACCTGGCTGC 204
Db	
OY	125 CCCATCTGTGAGCAGATAGCAGGAAGTCCCGACACTCTCCCGATATACAGACCCAGTCTGT 184
Db	
OY	205 GGCACGTAGTGGCTACATATATACGAATGAATGACAGCTCTGCTTGGCCCGGATAAAAAC 264
Db	
OY	185 GGCACAGATGGGGTACATACAGAGATGAATGCAAGCTCTGCTTGGCTGGATGAAAAC 244
OY	265 AAACAGGACATCCAGATATGAAGAATGGCAATGCTGATCCACAGAGACACTCAAGC 324
Db	
OY	245 AAACAGGACATCCAGATATGGAAGATGGCAAAATGCTGACTCCACAGAGAGCCCTTAAGC 304
OY	325 CATCAAGATGTC-ACCTGTGAGAAACAGTGGTGGCATGGAGAGATATGACATGAATAAA 383
Db	
OY	305 CATCAAGCTTGGGCTGGAGAGACAGCTCTGGGTGTGGAGAGATGTACATTAATAAAA 364
RESULT 10	
MMMPGC60	
LOCUS	515 bp mRNA Linear ROD 25-AUG-1997
DEFINITION	M.musculus mRNA for Mpgc60 protein.
ACCESSION	Y11505
VERSION	Y11505.1 GI:2344863
KEYWORDS	mpgc60 gene.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; 1 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Krause, R., Hemberger, M., Messerschmid, M., Kotthar, R. and Fündele, R. Molecular cDNA cloning and expression of murine Mpgc60, a gene predominantly expressed in the intestinal tract 2 (bases 1 to 515)
JOURNAL	
REFERENCE	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:52:25 ; Search time 274.744 Seconds
(without alignments)
3910.464 Million cell updates/sec

Title: US-09-941-309-2

Perfect score: 398
Sequence: 1 gcagcgcccgagccagctcag.....taaaagatccagcccaactg 398

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_19Jun03: *
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	398	100.0	398 18 AAT64673	Human PEC-60-type
2	398	100.0	402 20 AAT77490	Human ovarian tumor
3	395	99.2	433 21 AAC96176	Human colon cancer
4	395	99.2	433 22 AAH34805	Human colon cancer
5	388	97.5	389 19 AAV38073	Human tumour-assoc
6	349.4	87.8	394 24 ABV97573	Human pancreatic c
7	325.8	81.9	468 24 ABO60075	Human colon cancer
8	321.4	80.8	323 24 ABV86895	Human colon cancer

9	321.4	80.8	323 24 ABV89019	Human colon cancer
10	254.8	64.0	283 24 ABO60577	Human colon cancer
11	101.6	25.5	166 24 ABO65651	Human colon cancer
12	60	15.1	60 24 ABN39626	Human spliced tran
13	45.6	11.5	573 21 AAC04140	Human secreted pro
14	45.6	11.5	621 21 AAC04139	Human secreted pro
15	44.6	11.2	168 9 AAN80453	Sequence encoding
16	44.6	11.2	182 9 AAN80030	Synthetic DNA enco
17	44.6	11.2	182 9 AAN82216	Synthetic DNA enco
18	43	10.8	182 9 AAN82217	Synthetic DNA enco
19	43	10.8	182 9 AAN82232	Synthetic DNA enco
20	42.4	10.7	314 12 AAQ12109	Human pancreatic s
21	42.2	10.6	368 24 ABR86355	Human pancreas sec
22	42.2	10.6	368 24 ABN97352	Gene #3850 used to
23	42.2	10.6	368 24 ABL64781	Lung cancer relate
24	42.2	10.6	368 24 ABL65420	Lung cancer relate
25	42	10.6	182 9 AAN82227	Synthetic DNA enco
26	42	10.6	182 9 AAN82228	Synthetic DNA enco
27	42	10.6	182 9 AAN82233	Synthetic DNA enco
28	41.4	10.4	168 9 AAN80452	Sequence encoding
29	41.4	10.4	182 9 AAN82221	Synthetic DNA enco
30	41.4	10.4	415 17 AAT15982	PSKAB8 fragment #1
31	41.4	10.4	482 13 AAQ24083	Secretion plasmid
32	41.2	10.4	171 11 AAQ01494	Modified human pan
33	40.6	10.2	182 9 AAN82224	Synthetic DNA enco
34	40.6	10.2	182 9 AAN82225	Synthetic DNA enco
35	40.6	10.2	182 9 AAN82226	Synthetic DNA enco
36	40.6	10.2	182 9 AAN82223	Synthetic DNA enco
37	40.6	10.2	341 9 AAN80460	Human pancreatic s
38	40.6	10.2	362 24 ABO60575	Human colon cancer
39	40.6	10.2	39 24 ABO57663	Human signalling p
40	40.6	10.2	432 25 ACA56652	Human pancreatic c
41	40.6	10.2	604 21 AAC99064	Human pancreatic c
42	40.6	10.2	608 20 AA241396	Human normal pancr
43	40.6	10.2	692 23 ABV22245	Human prostate exp
44	40.6	10.2	692 23 ABV28082	Human prostate exp
45	40.6	10.2	904 23 ABV23510	Human prostate exp

ALIGNMENTS

RESULT 1	
ID AAT64673	standard; cDNA to mRNA; 398 BP.
AC AAT64673;	
XX	
XX	
XX	22-DEC-1997 (first entry)
XX	
XX	Human PEC-60-type protein encoding cDNA.
XX	
XX	Human; PEC-60-type protein; digestive disorder; immune disorder;
KW	neurous disorder; gastrointestinal tract hormone; gastric tissue;
KW	antigen; antibody; ds.
XX	
OS	Homo sapiens.
XX	
XX	
FT	Key
FT	CDS
FT	
FT	sig_peptide
FT	
FT	mat_peptide
FT	
XX	
XX	WO9715596-A1.
XX	
XX	01-MAY-1997.
XX	
XX	22-OCT-1996; 96WO-JP03061.
XX	

Location/Qualifiers
43..303
/*tag= a
/product= PEC-60-type-protein
/*tag= b
121..300
/*tag= c

PR 27-OCT-1995; 95JP-0280272.
 XX
 XX (SAGA) SAGAMI CHEM RES CENTRE.
 XX
 XX Kamata K, Kato S, Sekine S, Yamaguchi T.
 XX WPI: 1997-258958/23.
 DR P-PSDB: AAM15067.
 XX
 XX Human PEC-60-type protein - for treatment or diagnosis of digestive,
 PT immune or nervous disorders
 XX
 PS Claim 5; Page 10; 17pp; Japanese.
 CC The present sequence encodes a human PEC-60-type protein. The protein
 CC is a gastrointestinal tract hormone secreted by the gastric tissues.
 CC It is useful for treatment or diagnosis of digestive, immune or nervous
 CC disorders, or as an antigen to raise antibodies against them. Large
 CC amounts of the protein may be produced by using the corresponding cDNA.
 CC
 CC Sequence 398 BP; 107 A; 110 C; 105 G; 76 T; 0 other;
 QY
 Best Local Similarity 100.0%; Score 398; DB 18; Length 398;
 Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCAGGCCCCAGCAGCTCAGCTACACTATCCAGATCAGATGCGCCGCGCAGTGG 60
 Db 1 GCAGGCCCCAGCAGCTCAGCTACACTATCCAGATCAGATGCGCCGCGCAGTGG 60
 QY 61 GTATGCGCCGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 Db 61 GTATGCGCCGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 QY 121 GGAAGCTCCCTTCTCAGAGATGCCATCTGTGAGACAGATGAGAGTCCAACTGT 180
 Db 121 GGAAGCTCCCTTCTCAGAGATGCCATCTGTGAGACAGATGAGAGTCCAACTGT 180
 QY 181 TCCAGATGTCACCACTGCTGCTGGCAGTGTGAGTGGCTCACAATATGAGTGGCAG 240
 Db 181 TCCAGATGTCACCACTGCTGCTGGCAGTGTGAGTGGCTCACAATATGAGTGGCAG 240
 QY 241 CTCTGCTTGGCCGGATTAACCAACAGGACATCCAGATGAGATGGAATGCG 300
 Db 241 CTCTGCTTGGCCGGATTAACCAACAGGACATCCAGATGAGATGGAATGCG 300
 QY 301 TGAATCCACAGAGACCTCAGCCATGAGTGTCACTGAGAGAACTGGTGGCATGG 360
 Db 301 TGAATCCACAGAGACCTCAGCCATGAGTGTCACTGAGAGAACTGGTGGCATGG 360
 QY 361 AGAGGATATGACATGAATTAAGATCCAGCCCACTG 398
 Db 361 AGAGGATATGACATGAATTAAGATCCAGCCCACTG 398
 RESULT 2
 AA277490
 ID AA277490 standard; cDNA; 402 BP.
 XX
 XX AA277490;
 AC
 XX 10-Apr-2000 (first entry)
 DT
 XX Human ovarian tumor cDNA library derived EST fragment 41.
 DE
 XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;
 KW gene therapy; treatment; ss.
 XX
 OS Homo sapiens.
 XX
 PN DE19817557-A1.
 XX
 PD 21-OCT-1999.

XX
 PF 09-APR-1998; 98DE-1017557.
 XX
 XX 09-APR-1998; 98DE-1017557.
 PR
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 PI WPI: 1999-591920/51.
 DR P-PSDB: AAY76598, AAY76599.
 XX
 XX
 XX New nucleic acid sequences expressed in ovarian, and some other, cancer
 PT tissues, and derived polypeptides, for treatment of ovarian cancer and
 PT identification of therapeutic agents
 PS
 PS Claim 3; Page 170; 310pp; German.
 CC This invention describes novel nucleic acid (cDNA) sequences (A) which
 CC have anticancer activity and are highly expressed in ovarian tumor
 CC tissue (and some also in testis and breast cancer tissue). The products
 CC of the invention can be used for gene therapy. (A) are used (i) for
 CC recombinant expression of polypeptides (B) and (ii) to isolate complete
 CC genes. (B) are used (i) to identify agents suitable for treatment of
 CC ovarian cancer; (ii) directly for treating this form of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same unknown gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. AA277450-277572 represent the human
 CC ovarian tumor cDNA library derived EST fragments described in the method
 CC of the invention and encode the protein fragments represented in
 CC AAY76505-Y76638.
 CC
 CC Sequence 402 BP; 108 A; 111 C; 107 G; 76 T; 0 other;
 SO
 QY
 Best Local Similarity 100.0%; Score 398; DB 20; Length 402;
 Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCAGGCCCCAGCAGCTCAGCTACACTATCCAGATCAGATGCGCCGCGCAGTGG 60
 Db 4 GCAGGCCCCAGCAGCTCAGCTACACTATCCAGATCAGATGCGCCGCGCAGTGG 63
 QY 61 GTATGCGCCGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 Db 64 GTATGCGCCGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
 QY 121 GGAAGCTCCCTTCTCAGAGATGCCATCTGTGAGACAGATGAGAGTCCAACTGT 180
 Db 124 GGAAGCTCCCTTCTCAGAGATGCCATCTGTGAGACAGATGAGAGTCCAACTGT 183
 QY 181 TCCAGATGTCACCACTGCTGCTGGCAGTGTGAGTGGCTCACAATATGAGTGGCAG 240
 Db 184 TCCAGATGTCACCACTGCTGCTGGCAGTGTGAGTGGCTCACAATATGAGTGGCAG 243
 QY 241 CTCTGCTTGGCCGGATTAACCAACAGGACATCCAGATGAGATGGAATGCG 300
 Db 244 CTCTGCTTGGCCGGATTAACCAACAGGACATCCAGATGAGATGGAATGCG 303
 QY 301 TGAATCCACAGAGACCTCAGCCATGAGTGTCACTGAGAGAACTGGTGGCATGG 360
 Db 304 TGAATCCACAGAGACCTCAGCCATGAGTGTCACTGAGAGAACTGGTGGCATGG 363
 QY 361 AGAGGATATGACATGAATTAAGATCCAGCCCACTG 398
 Db 364 AGAGGATATGACATGAATTAAGATCCAGCCCACTG 401
 RESULT 3

ID	AAC98176	standard; CDNA: 433 BP.	
AC	AAC98176;		
XX			
DT	09-MAR-2001	(first entry)	
XX			
DE	Human colon cancer antigen nucleotide sequence SEQ ID NO:186.		
XX			
KM	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KM	identification; cytostatic; cardiocologic; neuroprotective; vulnerary;		
KM	immunomodulatory; muscular; gynaecological; gastrointestinal;		
KM	nephrotropic; antinfective; antibacterial; gene therapy; wound;		
KM	neural disorder; immune system disorder; muscular disorder;		
KM	reproductive disorder; gastrointestinal disorder; renal disorder;		
KM	infectious disease; cardiovascular disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200055351-A1.		
DD			
XX	21-SEP-2000.		
XX			
PE	08-MAR-2000; 2000WO-US05883.		
XX			
PR	12-MAR-1999; 99US-0124270.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI			
PI	Rosen CA, Ruben SM;		
XX			
DR	WPI: 2000-587534/55.		
DR	P-PSDB: AAB53419.		
XX			
PT	Colon cancer associated gene sequences, referred to as colon cancer		
PT	antigens, useful for the treatment, prevention, and diagnosis of colon		
PT	disorders such as colon cancer -		
XX			
PS	Claim 1; Page 610; 2104pp; English.		
XX			
CC	AAC97991 to AAC98763 encode the human colon cancer associated proteins,		
CC	called human colon cancer antigens, given in AAB53234 to AAB54006. The		
CC	human colon cancer antigens can have cytostatic, cardiocactive, muscular;		
CC	neuroprotective, immunomodulatory, gynaecological, gastrointestinal,		
CC	vulnerary, nephrotropic, antinfective and antibacterial activities, and		
CC	can be used in gene therapy. The colon cancer antigen polynucleotides,		
CC	proteins and antibodies to the proteins are useful for the prevention,		
CC	treatment and diagnosis of colon disorders, such as colon cancer. The		
CC	polynucleotides may be used in diagnostics and research, such as for		
CC	chromosome identification, and as hybridisation probes. The proteins		
CC	may also be used to prevent diseases such as neural disorders, immune		
CC	system disorders, muscular disorders, reproductive disorders,		
CC	gastrointestinal disorders, wounds, renal disorders, infectious		
CC	diseases, and cardiovascular disorders. AAC98764 to AAC98772 and		
CC	AAB54007 represent sequences used in the exemplification of the present		
CC	invention.		
XX			
XX			
SQ	Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;		
QY			
Query Match	99.2%; Score 395; DB 21; Length 433;		
Best Local Similarity	100.0%; Pred. No. 1.1e-114;		
Matches 395; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
DB			
17	GCAGGCCCCCAGCAGCTCAGGCTACATCCAGATCCAGATGAGCGCTCCGCCAGTGG	60	
17	GCAGGCCCCCAGCAGCTCAGGCTACATCCAGATCCAGATGAGCGCTCCGCCAGTGG	76	
QY	61	GTATCGCCCTGCGCTTGGCTCCCTCTTGTGTGGAGCAGGAGTGGCAGTGGCAGCA	120
DB	77	GTATCGCCCTGCGCTTGGCTCCCTCTTGTGTGGAGCAGGAGTGGCAGTGGCAGCA	136
QY	121	GGAAGCTCCCTTTTCAAGAATGCCCATCTGTGAACACATGTAAGTCTCAACCTGT	180

Db	137	GGAAAGCTCCCTTCTTCAGAAATGGCCCATCTGTGAACACATGTTAGAGTCTCCAACTGT	196
Qy	161	TCCCAGATGTCCAACTGGTCTGGCGGCACTGATGGGCTACATATACGAATGAATGCCAG	240
Db	197	TCCCAGATGTCCAACTGGTCTGGCGGCACTGATGGGCTACATATACGAATGAATGCCAG	256
Qy	241	CTCTGCTTGGCCCCGATATAAAACCAACGAGACATCCAGATCATGAAAGATGGCAATGC	300
Db	257	CTCTGCTTGGCCCCGATATAAAACCAACGAGACATCCAGATCATGAAAGATGGCAATGC	316
Qy	301	TGATCCCAAGAGAGCAACCTCAAGCCATGAAAGTCACTGGAGAACAGTGTGGCATGG	360
Db	317	TGATCCCAAGAGAGCAACCTCAAGCCATGAAAGTCACTGGAGAACAGTGTGGCATGG	376
Qy	361	AGAGGATATGACATGAAATMAAAGATCCAGCCCA	395
Db	377	AGAGGATATGACATGAAATMAAAGATCCAGCCCA	411
RESULT 4			
AAH34805	ID	AAH34805 standard; cDNA; 433 BP.	
XX	AAH34805;		
AC	03-SEP-2001	(first entry)	
XX			
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:1887.		
XX			
XX	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KM	colorectal carcinoma; ss..		
XX			
OS	Homo sapiens.		
XX			
PN	WO200122920-A2.		
XX			
PD	05-APR-2001.		
XX			
XX	28-SEP-2000; 2000WO-US26524.		
PF			
XX	29-SEP-1999; 99US-0157137.		
PR	03-NOV-1999; 99US-0163280.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX			
DR	WPI: 2001-235357/24.		
XX	P-PSDB: AAG75400.		
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,		
XX	useful for preventing, diagnosing and/or treating colorectal cancers -		
XX	Claim 1; Page 3403; 9803pp; English.		
PS			
XX	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon		
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where		
CC	the proteins are collectively known as colon cancer antigens. The colon		
CC	cancer antigens have cytostatic activity and can be used in gene		
CC	therapy and vaccine production. N and P may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate P		
CC	expression. For example, N and P may be used to treat disorders		
CC	associated with decreased expression by rectifying mutations or deletions		
CC	in a patient's genome that affect the activity of P by expressing		
CC	inactive proteins or to supplement the patients own production of P.		
CC	Additionally, N may be used to produce the colon cancer-associated P,		
CC	by inserting the nucleic acids into a host cell and culturing the cell		
CC	to express the proteins. N and P can be used in the prevention, diagnosis		
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204		
CC	and AAB7789 represent sequences used in the exemplification of the		
CC	present invention.		
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were		
CC	missing at time of publication, meaning no sequences are present for		

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

SQ Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;

Query Match	99.28;	Score 395;	DB 22;	Length 433;
Best Local Similarity	100.08;	Pred. No. 1.7e-114;		
Matches 395;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	GCAGGCCCCAGGACAGCTCAGGGCTACATCATCCAGAGATACCATAGGGCCGTCGCAATGG	60
Db	17	GCAGGCCCCAGGACAGCTCAGGGCTACATCATCCAGAGATACCATAGGGCCGTCGCAATGG	76
QY	61	GTATCGCCCTGAGCTTGGCTGCCCCCTCTTGTGTGGACAGGAGGAGCCAGTGGACGA	120
Db	77	GTATCGCCCTGAGCTTGGCTGCCCCCTCTTGTGTGGACAGGAGGAGCCAGTGGACGA	136
QY	121	GGAAAGCTCCCTTTCTCAAGAAATGCCATCTGTGTGAACACATGTAGAGTCTCCAACTGT	180
Db	137	GGAAAGCTCCCTTTCTCAAGAAATGCCATCTGTGTGAACACATGTAGAGTCTCCAACTGT	196
QY	181	TCCCAATGTCCAACTGTGTCGGGGCACTATATGGGCTCACATATACGAATGAATGGCAG	240
Db	197	TCCCAATGTCCAACTGTGTCGGGGCACTATATGGGCTCACATATACGAATGAATGGCAG	256
QY	241	CTCTGCTTGCCCGGATAAAAACAACAGACATCCGATCATGAAGAATGGCAATGC	300
Db	257	CTCTGCTTGCCCGGATAAAAACAACAGACATCCGATCATGAAGAATGGCAATGC	316
QY	301	TGATCCCAAGAGACACCTCAAGCCCATGAAGTGTACGCTGGAGAAACAGTGGTGGCATGG	360
Db	317	TGATCCCAAGAGACACCTCAAGCCCATGAAGTGTACGCTGGAGAAACAGTGGTGGCATGG	376
QY	361	AGAGGATATGACATGAATATAAAGATCCAGCCCA	395
Db	377	AGAGGATATGACATGAATATAAAGATCCAGCCCA	411

RESULT 5
AAV38073

ID	AAV38073	standard; CDNA; 389 BP.
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

AC AAV38073;

DT 14-SEP-1998 (first entry)

Human tumour-associated Kazal inhibitor encoding cDNA

Human; tumour-associated Kazal inhibitor; TAK1; inflammation; cancer; ulcerative colitis; Crohn's disease; inflammatory cytokine; elastase; serine protease; glycoprotein hormone; trypsin; chymotrypsin; pepsin; ss

OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	CDS	34..294

FT	/product= "tumour-associated Kazal inhibitor"
XX	
PN	WO9820132-A1.

PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Goli SK, Murry LE,

DR WPI; 1998-286948/25.

XX

XX Claim 5, Fig 1; 62pp. English.

CC The present sequence encodes human tumour-associated Kazal inhibitor
CC (TAKI). TAKI is expressed in response to inflammatory cytokines and
CC functions in the inhibition of serine proteases or glycoprotein hormone
CC present in diseased tissues. The TAKI can be used to inhibit the tissue
CC destruction associated with the production of excess proteases such as
CC trypsin, chymotrypsin, elastase, or pepsin. Products from the present
CC invention can be used in the treatment of e.g. proliferative cell
CC division in inflamed intestinal tissues, ulcerative colitis, Crohn's
CC disease or cancer. The products can also be used for detection and
CC diagnosis.

SQ Sequence 389 BP; 105 A; 105 C; 102 G; 76 T; 1 other;

Query Match	97.5%	Score 388	DB 19	Length 389
Best Local Similarity	100.0%	Pred. No.	1.7e-112	
Matches 388	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0

QY	11	GCAGCTCAGGCTACACTATCCAGAGTACAGATGCGCGTCCCGCACTGGGTAAATGCCCC	70
Db	2	GCAGGTCAAGGCTACACTATCCAGAGTACAGATGCGCGTCCCGCAGTGGGTAAATGCCCC	61
QY	71	TGGCCTTGGCGTCCCTCTTGTGTGTGGACAGGGAAGTCCAGTGGACAGCAAGAAAGCTCC	130
Db	62	TGGCCTTGGCGTCCCTCTTGTGTGTGTGGACAGGGAAGTCCAGTGGACAGCAAGAAAGCTCC	121
QY	131	CTTTCTCAAGAAATGGCCATCTGTGAACAACATGGTAGACTCCAAACCTGTTTCCAGATGT	190
Db	122	CTTTCTCAAGAAATGCCCATCTGTGAACAACATGGTAGACTCTCAACCTGTTTCCAGATGT	181
QY	191	CCAACTGGCTGTGCGGCACTGATGGGGCTCACATATAGCAATGAATGCCAGCTGTGCTTGG	250
Db	182	CCAACTGGCTGTGCGGCACTGATGGGGCTCACATATAGCAATGAATGCCAGCTGTGCTTGG	241
QY	251	CCCGGATAAAAACCAAAACAGACATCCAGATCATTAAGATGGGCAAAATGCTGATCCCA	310
Db	242	CCCGGATAAAAACCAAAACAGACATCCAGATCATTAAGATGGGCAAAATGCTGATCCCA	301
QY	311	GGAGCACTCAAGCCATGAAGTGTCAAGCTGGAGAAACACTGTTGGGCACTGGAGAGATATG	370
Db	302	GGAGCACTCAAGCCATGAAGTGTCAAGCTGGAGAAACACTGTTGGGCACTGGAGAGATATG	361
QY	371	ACATGAATATAAAGATCCAGCCCAACTG	398
Db	362	ACATGAATATAAAGATCCAGCCCAACTG	389

RESULT 6

ID	ABV97573	standard; cDNA; 394 BP

AC ABV97573;

DT 14-JAN-2003 (first entry)

DE Human pancreatic cancer expressed cDNA SEQ ID NO 2981.

KW Human, pancreas; cancer; gene therapy; vaccine; immunostimulant.
KW Cytostatic; tumour; gene; ss.

OS Homo sapiens

PN WO200260317-A2.

PD 08-AUG-2002.

PF 30-JAN-2002; 2002WO-US02781.

QY	240	GCCTCTGCTTGGCCCCGGATATAAAACCAACAGACATCC	AGATCATGAAGAGTGGCAAT	2368
Db	247	NCCTGCTGTTGGCCCCGGATATAAAACCAACAGACATCC	AGATCATGAAGAGTGGCAAT	3068
QY	299	GCATGATCCACAGAGACACCTCAACCCATGAAGTGTACGTG	GAGAACAGTGTGGGCAT	3588
Db	307	GCATGATCCACCAAGACACCTCAACCCATGAAGTGTACGTG	GAGAACAGTGTGGGCAT	3668
QY	359	GGAGA---GGATGTGCATGAAATTAAGAATCCACCCAACT		397
Db	367	TNGAANAGGATTGTGCATGAAATTAAGAATCCACCCAACT		408

RESULT	ID	ABV86895/c	standard; cDNA; 323 BP.
XX	ABV86895;		
XX	13-DEC-2002	(first entry)	
XX	Human colon cancer related	cDNA SEQ ID NO 206.	

KW Human, colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene
 KM ss.
 KN
 XX Homo sapiens.
 OS
 XX W0200258534-A2.
 PN
 XX
 XX
 PD 01-AUG-2002.
 XX
 XX
 PF 19-NOV-2001; 2001WO-US43704.
 XX
 PR 20-NOV-2000; 2000US-252222P.
 PR 06-FEB-2001; 2001US-267011P.
 PR 28-MAR-2001; 2001US-279670P.
 PR 10-JUL-2001; 2001US-304037P.
 XX
 XX (CORI-) CORIYA CORP.

PI Stolk JA, Xu J, Chénault RA, Meagher MJ, Secrist H, King GE,
XX
DR WPI; 2002-608400/65.

Claim 1; SEQ ID NO 206; 266pp + Sequence Listing; English.

PS Claim 1, SEQ ID NO 206; 266pp + Sequence Listing; English.

XX

CC The invention relates to a human colon tumour expressed polynucleotide
CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (I) any
CC 2600 fully defined nucleotide sequences (ABW6659-ABW69289); (II)
CC complements of (I); (III) at least 20 continuous residues of (I); (IV)
CC sequences that hybridize to (I), under moderately stringent conditions;
CC (V) sequences having at least 75% or 90% identity to (I); or (VI)
CC degenerate variants of (I). The compositions and methods of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC cancer, particularly colon cancer. (I) can be used in gene therapy and
CC (I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

SQ Sequence 323 BP; 66 A; 81 C; 93 G; 83 T; 0 other;

Query Match	80.8%	Score 321.4;	DB 24;	Length 323;
Best Local Similarly	99.7%	Pred. No. 1.8e-91;		
Matches 322; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

15 GCTCAGGCTACACTATCCGAGATCAGCATGGCCGTCGCCAGTGGTATCGCCCTGGC 74

Db	323	GCCTCAGGCTACACTATCCCGAGATCAGCATGGCCGTCGCCACGTGGATTAATACGCCCTCGGC	264
OY	75	CTTGGCTGCCCTCCTTGTGTGGACAGGAACTGCCATGGCAGCAGGAAAGCTCCCTTT	134
Db	263	CTTGGCTGCCCTCCTTGTGTGGACAGGAAAGTGCACATGGCAGCAGGAAAGCTCCCTT	204
OY	135	CTCAAGAAATGCCCATCTGTGTGAACACATGCTAGTCAAGTCTCCAACTCTTCCCGATGTCCAA	194
Db	203	CTCAAGAAATGCCCATCTGTGTGAACACATGCTAGTCAAGTCTCCAACTCTTCCCGATGTCCAA	144
OY	195	CCTGGTCTGGGGCAGCTGATGGGCTCACATATATCGAATGAATGCCAGCTCTGTTGGCCG	254
Db	143	CTGTGCTGGGGCAGCTGATGGGCTCACATATACGAATGAATGCCAGCTCTGTTGGCCG	84
OY	255	GATAAAAAACCAACAGGACATCCAGATCATGAAAGATGGCAAAATGCTGATCCACAGGAG	314
Db	83	GATAAAAAACCAACAGGACATCCAGATCATGAAAGATGGCAAAATGCTGATCCACAGGAG	24
OY	315	CACCTCAAGCCATGAAGTGTCAAG	337
Db	23	CACCTCAAGCCATGAAGTGTCAAG	1

RESULT 9	
ABV89019	
ID	ABV89019 standard; cDNA; 323 BP.

AC ABV89019;

DT 13-DEC-2002 (first entry)

DE	Human colon cancer related	CDNA SEQ ID NO	2330
DE	Human colon cancer related	CDNA SEQ ID NO	2330

KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene
KW ss.

OS Homo sapiens.

PN W0200258534-A2

PD 01-AUG-2002.

19-NOV-2001; 2001WO-US43704.

20-NOV-2000; 2000US-252222P.

PR 28-MAR-2001; 2001US-279670P.

[illegible]

XX (CONT.) CONTAIN CONT

PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secríst H, King GE;

DR WPI; 2002-608400/65.

PT New isolated tumor colon polynucleotide and polypeptide, useful for the

[illegible]

PS Claim 1; SEQ ID NO 2330; 266pp + Sequence Listing; English.

CC The invention relates to a human colon tumour expressed polynucleotide

2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)

CC sequences that hybridize to (i), under moderately stringent conditions;

degenerate variants of 'i). The compositions and methods of the present

CC cancer, particularly colon cancer. (I) can be used in gene therapy and

CC Note: The sequence data for this patent did not form part of the printed

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 11:16:55 : Search time 2358.39 seconds
(without alignments)
4101.602 Million cell updates/sec

Title: US-09-941-309-2
Perfect score: 398
Sequence: 1 gcacgagcccccagcagctcag.....taaaagatccagcccaactg 398

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	398	100.0	553	10	AM960685 EST372756
2	355	89.2	355	12	BM856263 K-EST0139
3	330.8	83.1	334	13	BM636767 he01b10.y
4	307.4	77.2	321	9	AA534438 nfb0b04.s

5	283.4	71.2	1115	10	BE543478
6	282	70.9	537	10	BG207464
7	227	57.0	232	7	AA296922
8	211.2	53.1	420	4	BS12489
9	207.2	52.1	220	9	AT732343
10	194	48.7	352	9	AA823442
11	190.6	47.9	364	13	BY089445
12	186.2	46.8	318	10	BE136647
13	185.2	46.5	330	9	AA692430
14	184.2	46.3	325	14	CB222030
15	184	46.2	395	9	AV066321
16	182.8	45.9	390	13	BY454425
17	182.2	45.8	361	9	AA591318
18	178.2	44.8	313	9	AA689871
19	177	44.5	210	12	BM856250
20	172.2	43.3	285	9	AA982819
21	170.2	42.8	316	9	AA821396
22	169.6	42.6	350	9	AA277286
23	167.8	42.2	433	9	AI009059
24	166	41.7	300	9	AA710853
25	163.4	41.1	370	9	AV066728
26	161.6	40.6	300	9	AV067091
27	161	40.5	293	12	BM430399
28	160.2	40.3	277	9	AV052389
29	153.2	38.5	277	9	AV067355
30	152.4	38.3	287	9	AV070840
31	150.6	37.8	369	9	AA733374
32	145.8	36.6	277	9	AV068882
33	145.8	36.6	582	9	AA502919
34	144.4	36.3	146	10	BF756812
35	144.2	36.2	542	9	AT732203
36	140.4	35.3	266	9	AV062346
37	137.8	34.6	242	9	AV066442
38	135.2	34.0	220	10	BE138001
39	135.2	34.0	229	9	AA592368
40	134.2	33.7	268	9	AA512376
41	130.8	32.9	214	10	BE199860
42	127.4	32.0	282	9	AV063285
43	127.4	32.0	286	9	AV064715
44	126.4	31.8	131	10	BG202409
45	119.4	30.0	301	9	AV068336

ALIGNMENTS

RESULT 1
LOCUS AM960685 553 bp mRNA linear EST 01-JUN-2000
DEFINITION EST372756 MAGE resequences, MAGE Homo sapiens CDNA, mRNA sequence.
ACCESSION AM960685
VERSION AM960685.1 GI:8150369
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE Hegde, P., Ol, R., Abernathy, K., Dharp, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
ASSESSMENT OF gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

TITLE JOURNAL COMMENT
Unpublished
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850-USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@igf.org
Plate: 149
Seq primer: Reverse.
Location/Qualifiers

FEATURES

	source	1..553	/organism="Homo sapiens"	
			/mol_type="mRNA"	
			/db_xref="taxon:9606"	
			/clone_id="MAGE_resequences, MAGE"	
			/note="Vector: pBuescriptsm"	
BASE COUNT	156 a	153 c	137 g	107 t
ORIGIN				
Query Match	100.0%	Score 398:	DB 10;	Length 553;
Best Local Similarity	100.0%;	Pred. No. 1.le-101;		
Matches 398:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1	GCAGGCCCCCAGCACCCTACAGGTACATCCAGATCCAGATGGCGTCCGCACTGG 60		
Db	3	GCAGGCCCCCAGCACCCTACAGGTACATCCAGATCCAGATGGCGTCCGCACTGG 62		
OY	61	GTAATCGCCCTGGCCCTTGTCGCCCTCTTTGTGTGAACAAGAAAGTCCAGTGGCACCA 120		
Db	63	GTAATCGCCCTGGCCCTTGTCGCCCTCTTTGTGTGAACAAGAAAGTCCAGTGGCACCA 122		
OY	121	GGAAAGCTCCCTTTTCTCAAGAATGCCCATCTGTGTGAACAATGTTAATCTCCAACTGT 180		
Db	123	GGAAAGCTCCCTTTTCTCAAGAATGCCCATCTGTGTGAACAATGTTAATCTCCAACTGT 182		
OY	181	TCCCAGATGTCCAACCTGTGTCGGGGCCTGATGGGGCTACATATACGAATGAATGCCAG 240		
Db	183	TCCCAGATGTCCAACCTGTGTCGGGGCCTGATGGGGCTACATATACGAATGAATGCCAG 242		
OY	241	CTCTGCTTGGCCCGGATAAAAAACCAACAGACATCCAGATCATGAAGATGGCAAATGC 300		
Db	243	CTCTGCTTGGCCCGGATAAAAAACCAACAGACATCCAGATCATGAAGATGGCAAATGC 302		
OY	301	TGATCCACACAGAGACCTCAAGCCATGAAGTGTCAAGCTGGAGAACAGTGGTGGCATATG 360		
Db	303	TGATCCACACAGAGACCTCAAGCCATGAAGTGTCAAGCTGGAGAACAGTGGTGGCATATG 362		
OY	361	AGAGGATATGACATGAATAAAGATCCAGCCCAACTG 398		
Db	363	AGAGGATATGACATGAATAAAGATCCAGCCCAACTG 400		
RESULT 2				
BM856263		355 bp	mRNA	linear EST 06-MAR-2002
LOCUS	K-EST0139933 S14K402 Homo sapiens cDNA clone S14K402-51-E10 5'			
DEFINITION	K-EST0139933 S14K402 Homo sapiens cDNA clone S14K402-51-E10 5'			
ACCESSION	BM856263			
VERSION	BM856263.1 GI:19212662			
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	1 (bases to 355)			
	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,			
	Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and			
	Kim,Y.S.			
	21C Frontler Korean EST Project 2001			
TITLE	Unpublished			
JOURNAL	Contact: kim ys			
COMMENT	Genome Research Center			
	Korea Research Institute of Bioscience & Biotechnology			
	52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea			
	Tel.: +82-42-860-4470			
	Fax: +82-42-860-4409			
	Email: yongsungemil.kribb.re.kr			
	Plate: 51 row: E column: 10			
	High quality sequence stop: 355.			
	location/Qualifiers			
	1..355			
FEATURES	/organism="Homo sapiens"			
source	/mol_type="mRNA"			

	/db_xref="taxon:9606"	
	/clone="S14K402-51-E10"	
	/cell_line="R402"	
	/lab_host="Top10F"	
	/clone_id="S14K402"	
	/note="Organ: Stomach; Vector: pTZ18BPI; Site.1: EcoRI; Site.2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with df-tailed vector. The df-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."	

BASE COUNT	90 a	102 c	94 g	69 t
ORIGIN				

Query Match	89.2%;	Score 355;	DB 12;	Length 355;
Best Local Similarity	100.0%;	Pred. No. 1.3e-89;		
Matches 355:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	3	AGGCCCGCAGCGCTACACTATCCGAGATCAGCATGGCCGCAGTGAGGT 62
Db	1	AGGCCCGCAGCGCTACACTATCCGAGATCAGCATGGCCGCAGTGAGGT 60
OY	63	AATGGCCCTGGCCTTTGGCTGCCCTTCCTTTGTTGTGGACAGGGAATGCCAGTGACAGAG 122
Db	61	AATGGCCCTGGCCTTTGGCTGCCCTTCCTTTGTTGTGGACAGGGAATGCCAGTGACAGAG 120
OY	123	AAAGTCCTCTTCAAGAATGCCATCTGTGAACAACATGGTAGTATCCCAACCTGTC 182
Db	121	AAAGTCCTCTTCAAGAATGCCATCTGTGAACAACATGGTAGTATCCCAACCTGTC 180
OY	183	CCAGATGTCCAACCTGTGTGCGGACACTGATGAGGCTCACATATACGATGCAATGCCAGCT 242
Db	181	CCAGATGTCCAACCTGTGTGCGGACACTGATGAGGCTCACATATACGATGCAATGCCAGCT 240
OY	243	CTGCTTGGCCCGGATAAAAACCACACAGGACATCCAGATCATGAAAGATGCCAATGCTG 302
Db	241	CTGCTTGGCCCGGATAAAAACCACACAGGACATCCAGATCATGAAAGATGCCAATGCTG 300
OY	303	ATCCACAGGAGCACTCCAACGCATVGAAGTGCAGCTGGAACAAACAGTGGTGGA 357
Db	301	ATCCACAGGAGCACTCCAACCATGAGCTGCAGCTGGAGAACAGTGGTGGA 355

RESULT 3	
B0636767	
LOCUS	B0636767 334 bp mRNA linear EST 15-JUL-2002
DEFINITION	he01b10.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
ACCESSION	B0636767
VERSION	B0636767.1 GI:21761226
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman J.M., Bouffard,G., Smith,D. and Peterson,K. Expressed sequence tag analysis of human retina for the NEIBank Project: Rebinding, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts
JOURNAL	Mol. Vis. 8 (4), (2002) In press
COMMENT	Contact: Wistow G

Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graham@helix.nih.gov
Plate: 01 row: b column: 10
Seq primer: M13p1 reverse primer (ABI).

FEATURES

source

Location/Qualifiers

1..334

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="he01b10"

/tissue_type="Retina"

/dev_stage="Adult"

/lab_host="EMD10B"

/clone_lib="Human Retina cDNA (un-normalized, unamplified)

); hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCGTGATCGGCGGCCGC(7)15-3'] . EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT

94 a 86 c 88 g 66 t

ORIGIN

Query Match

Best Local Similarity 83.1%; Score 330.8; DB 13; Length 334;

Matches 332; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 59 GGGTAATCGCCCTGGCTGGCTGCTGCTGTTGTGTGAGAGGAAAGTCCAGTGGCAG 118
|||
DB 1 GGACAATCGCCCTGGCTGGCTGCTGCTGTTGTGTGAGAGGAAAGTCCAGTGGCAG 60
OY 119 CAGGAAGCTCCCTTTTCAGAGATGCCATCTGTGAACATGTGTAGAGTCTCCACCT 178
|||
DB 61 CAGGAAGCTCCCTTTTCAGAGATGCCATCTGTGAACATGTGTAGAGTCTCCACCT 120
OY 179 GTTCCAGATGTCCACCTGCTGCGGCACTGATGAGGCTCACAATATGCAATGAATGCC 238
|||
DB 121 GTTCCAGATGTCCACCTGCTGCGGCACTGATGAGGCTCACAATATGCAATGAATGCC 180
OY 239 AGCTCTGCTGGCCCGGATAAAACCAACAGACATCCAGATCATGAAGATGGCAAT 298
|||
DB 181 AGCTCTGCTGGCCCGGATAAAACCAACAGACATCCAGATCATGAAGATGGCAAT 240
OY 299 GCTATCCACAGAGAGCCTCAAGCATGAGTGTGAGGAGAGAGTGGGAT 358
|||
DB 241 GCTATCCACAGAGAGCCTCAAGCATGAGTGTGAGGAGAGAGTGGGAT 300
OY 359 GGAGAGATATGACATGAATAAAGATCCAGCC 392
|||
DB 301 GGAGAGATATGACATGAATAAAGATCCAGCC 334

RESULT 4

AA534438/c

LOCUS

AA534438 321 bp mRNA linear EST 21-Aug-1997

DEFINITION

n880b04.s1 NCI CGAP Co3 Homo sapiens cDNA clone IMAGE:926191 3'

similar to SW:P860_PIG P37109 PEPTIDE PEC-60 PRECURSOR. ; mRNA

sequence.

ACCESSION

AA534438

KEYWORDS

AA534438.1 GI:2278691

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

REFERENCE

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

COMMENT

Tumor Gene Index

UNPUBLISHED

Unpublished

CONTACT

Contact: Robert Strausberg, Ph.D.

EMAIL

Email: c9apb5-remail.nih.gov

TISSUE

Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA

DNA Sequencing by: Greg Lennon, Ph.D.

CLONE

Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at:

WWW

www.bio.illnl.gov/bdbr/image/image.html

INSERT

Insert length: 417 Std Error: 0.00

SEQ

Seq primer: -40m13 fwd. ET from Amer sham.

FEATURES

source

Location/Qualifiers

1..321

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:926191"

/sex="pooled"

/tissue_type="colon"

/lab_host="DH10B"

/clone_lib="NCI CGAP Co3"

/note="Vector: pT73B-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization."

BASE COUNT

62 a 83 c 81 g 95 t

ORIGIN

Query Match

Best Local Similarity 77.2%; Score 307.4; DB 9; Length 321;

Matches 319; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 79 GCTGCCCTCTTGTGTGAGAGGAAAGTCCAGTGGCAGAGGAAAGCTCCCTTTCGA 138
|||
DB 321 GCTGCCCTCTTGTGTGAGAGGAAAGTCCAGTGGCAGAGGAAAGCTCCCTTTCGA 262
OY 139 AGAATGCCATCTGTGAACACATGGTAGAGTCTCCACCTGTTCCAGATGCCAACCTG 198
|||
DB 261 AGAATGCCATCTGTGAACACATGGTAGAGTCTCCACCTGTTCCAGATGCCAACCTG 202
OY 199 GTTCGGGCACTGATGGGCTCACAATATGCAATG-AATGCCAGCTGCTGGGCCGAT 257
|||
DB 201 GTTCGGGCACTGATGGGCTCACAATATGCAATGCAAGCTCTGCTGGGCCGAT 142
OY 258 AAAAACCACAGAGACATCCAGATCATGAAGATGGCAAAATGCTGATCCACAGAGCAC 317
|||
DB 141 AAAAACCACAGAGACATCCAGATCATGAAGATGGCAAAATGCTGATCCACAGAGCAC 82
OY 318 CTCAAGCCATGAAGTCAAGTGGAGAACACTGTTGGGATGAGAGATATGACATGAA 377
|||
DB 81 CTCAAGCCATGAAGTCAAGTGGAGAACACTGTTGGGATGAGAGATATGACATGAA 22
OY 378 ATAAAGATCCAGCCCAACTG 398
|||
DB 21 ATAAAGATCCAGCCCAACTG 1

RESULT 5

BE543478

LOCUS BE534378 1115 bp mRNA linear EST 09-AUG-2000
 DEFINITION 601070792F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456839 5',
 mRNA sequence.
 ACCESSION BE534378
 VERSION BE534378.1 GI:9772123
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 1115)
 NIH-MGC http://mhc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 UNPUBLISHED
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.llnl.gov
 Plate: LLM8445 row: h column: 24
 High quality sequence stop: 385.
 Location/Qualifiers
 1..1115
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3456839"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_12"
 /note="Organ: cervix; Vector: pCMV-Sport6; Site:1; NotI;
 Site:2; SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."
 BASE COUNT 480 a 300 c 247 g 88 t
 ORIGIN
 Query Match 71.2%; Score 283.4; DB 10; Length 1115;
 Best Local Similarity 99.3%; Pred. No. 3.5e-69;
 Matches 295; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 103 GAAGTCCAGTGGCAGCAGGAAGCTCCCTTCTCAAGATGCCATCTGTGAACATG 162
 54 GAAGTCCAGTGGCAGCAGGAAGCTCCCTTCTCAAGATGCCATCTGTGAACATG 113
 163 GTAGAGTCTCCAACTGTTCCAGATGTCACACTGTGGGCACTGATGGGCTCACA 222
 114 GTAGAGTCTCCAACTGTTCCAGATGTCACACTGTGGGCACTGATGGGCTCACA 173
 QY 223 TATACGAATGAATGCCAGCTCTGTCGCGCGGATATAAACCAGACATCCAGATC 282
 174 TATACGAATGAATGCCAGCTCTGTCGCGCGGATATAAACCAGACATCCAGATC 233
 QY 283 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 342
 234 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 293
 QY 343 GAACAGTGTGGG-CATGAGAGAGATATGACATGAATAAAGATCCAGCCCACTG 398
 294 GAACAGTGTGGGCTTGGAGAGAGATATGACATGAATAAAGATCCAGCCCACTG 350
 Db
 RESULT 6
 BG207464 537 bp mRNA linear EST 21-APR-2001
 LOCUS RST26945 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG207464
 ACCESSION BG207464.1 GI:13729151
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 537)
 Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Leventhal, C., Thornton, M., Ramchandran, R., Whittington, J.,
 Lerner, L., Costanzo, D., McEligott, K., Booser, S., Mays, R., Smith,
 E., Veloso, N., Kika, A., Hess, J., Cochren, K., Lo, R., Offenbacher,
 J., Danzig, J., and Ducar, M.
 Creation of genome-wide protein expression libraries using random
 activation of gene expression
 Nat. Biotechnol. 19 (5), 440-445 (2001)
 TITLE 21227151
 JOURNAL PUBMED
 MEDLINE 11329013
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave., Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@athersys.com
 High quality sequence stop: 482.
 Location/Qualifiers
 1..537
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is Hri080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in Hri080 under normal circumstances."
 BASE COUNT 150 a 133 c 148 g 106 t
 ORIGIN
 Query Match 70.9%; Score 282; DB 10; Length 537;
 Best Local Similarity 100.0%; Pred. No. 6.6e-69;
 Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 103 GAAGTCCAGTGGCAGCAGGAAGCTCCCTTCTCAAGATGCCATCTGTGAACATG 162
 252 GAAGTCCAGTGGCAGCAGGAAGCTCCCTTCTCAAGATGCCATCTGTGAACATG 311
 QY 163 GTAGAGTCTCCAACTGTTCCAGATGTCACACTGTGGGCACTGATGGGCTCACA 222
 312 GTAGAGTCTCCAACTGTTCCAGATGTCACACTGTGGGCACTGATGGGCTCACA 371
 QY 223 TATACGAATGAATGCCAGCTCTGTCGCGCGGATATAAACCAGACATCCAGATC 282
 372 TATACGAATGAATGCCAGCTCTGTCGCGCGGATATAAACCAGACATCCAGATC 431
 QY 283 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 342
 432 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 491
 QY 343 GAACAGTGTGGGATGAGAGATATGACATGAATAAAG 384
 492 GAACAGTGTGGGATGAGAGATATGACATGAATAAAG 533
 Db
 RESULT 7
 AA296922 232 bp mRNA linear EST 18-APR-1997
 LOCUS EST112478 Retina II Homo sapiens cDNA 5' end similar to similar to
 DEFINITION AA296922
 ACCESSION AA296922
 VERSION AA296922.1 GI:1949254
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 11:18:55 : Search time 73.6762 Seconds
(without alignments)
2384.294 Million cell updates/sec

Title: US-09-941-309-2

Perfect score: 398

Sequence: 1 gcagcccccagccagctcag.....taaaagatccagccactg 398

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued Patents_NA: *
2: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	398	100.0	398	US-09-065-019-2	Sequence 2, Appli
2	388	97.5	388	US-08-744-670-2	Sequence 2, Appli
3	388	97.5	388	US-09-149-933-2	Sequence 2, Appli
4	180	45.2	180	US-09-065-019-1	Sequence 1, Appli
5	41.4	10.4	458	US-08-880-829-21	Sequence 21, Appli
6	40.6	10.2	432	US-08-235-515A-26	Sequence 26, Appli
7	39.6	9.9	302	US-08-221-298-56	Sequence 56, Appli
8	39.4	9.9	420	US-09-016-434-361	Sequence 361, App
9	37.4	9.4	285	US-09-107-532A-63	Sequence 63, Appli
10	36.6	9.2	1632	US-08-839-709-2	Sequence 2, Appli
11	33.4	8.4	309	US-09-204-859-2	Sequence 2, Appli
12	33.4	8.4	4112	US-08-340-203A-2	Sequence 2, Appli
13	33.4	8.4	4112	US-08-452-567-2	Sequence 2, Appli
14	33.4	8.4	4112	US-08-452-427-2	Sequence 2, Appli
15	33.4	8.4	4112	US-09-085-407-2	Sequence 2, Appli
16	33.4	8.4	4112	US-08-340-203A-1	Sequence 1, Appli
17	33.4	8.4	4616	US-08-452-567-1	Sequence 1, Appli
18	33.4	8.4	4616	US-08-452-427-1	Sequence 1, Appli
19	33.4	8.4	4616	US-09-085-407-1	Sequence 1, Appli
20	33.4	8.4	4616	US-09-085-407-1	Sequence 1, Appli
21	32.8	8.2	1959	US-09-528-959-2	Sequence 2, Appli
22	32.8	8.2	1959	US-09-528-959-2	Sequence 2, Appli
23	32.2	8.1	265	US-09-221-298-55	Sequence 55, Appli
24	31.8	8.0	423	US-09-252-991A-14360	Sequence 14360, A
25	31.8	8.0	1302	US-09-252-991A-14317	Sequence 14317, A
26	31.8	8.0	1479	US-09-252-991A-14286	Sequence 14286, A
27	31.8	8.0	1716	US-09-252-991A-14386	Sequence 14386, A

28	31.4	7.9	1129	US-09-383-586-10	Sequence 10, Appli
29	31.4	7.9	1206	US-09-312-283C-357	Sequence 357, App
30	31.2	7.8	3804	US-09-620-312D-894	Sequence 894, App
31	31	7.8	126	US-08-586-676E-4	Sequence 4, Appli
32	31	7.8	129	US-08-586-676E-5	Sequence 5, Appli
33	31	7.8	138	US-08-586-676E-6	Sequence 6, Appli
34	30.8	7.7	7218	US-08-232-463-14	Sequence 14, Appli
35	30.6	7.7	2972	US-08-720-484A-3	Sequence 3, Appli
36	30.6	7.7	2972	US-08-953-823A-3	Sequence 3, Appli
37	30.6	7.7	2972	US-09-398-239-3	Sequence 3, Appli
38	30.6	7.7	2972	US-09-560-876A-3	Sequence 3, Appli
39	30.6	7.7	3112	US-09-318-448-9	Sequence 3, Appli
40	30.4	7.6	48974	US-08-920-422-17	Sequence 17, Appli
41	29.8	7.5	463	US-09-439-313-362	Sequence 362, App
42	29.8	7.5	463	US-09-352-616A-362	Sequence 362, App
43	29.8	7.5	1290	US-09-328-352-2156	Sequence 2156, Ap
44	29.6	7.4	789	US-09-252-991A-2208	Sequence 2208, Ap
45	29.6	7.4	1059	US-09-252-991A-2370	Sequence 2370, Ap

ALIGNMENTS

RESULT 1	US-09-065-019-2	Sequence 2, Application US/09065019
Patent No. 6280968		
GENERAL INFORMATION:		
APPLICANT: Kato, Seishu		
APPLICANT: Yamaguchi, Tomoko		
APPLICANT: Sekine, Shingo		
APPLICANT: Kamata, Kouji		
TITLE OF INVENTION: HUMAN PEC-60-LIKE PROTEIN AND DNA ENCODING THIS PROTEIN		
FILE REFERENCE: 6700PCT-US		
CURRENT APPLICATION NUMBER: US/09/065,019		
CURRENT FILING DATE: 1998-04-17		
NUMBER OF SEQ ID NOS: 3		
SOFTWARE: Patent In Ver. 2.0		
SEQ ID NO 2		
LENGTH: 398		
TYPE: DNA		
ORGANISM: Homo sapiens		
US-09-065-019-2		
Query Match	100.0%: Score 398; DB 3; Length 398;	
Best Local Similarity	100.0%: Pred. No. 4.5e-121;	
Matches 398; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GCAGGCCCCAGCCAGCTAGGCTACATATCCAGATCAGCATGCCGTCGCCAGTGG 60	
DB	1 GCAGGCCCCAGCCAGCTAGGCTACATATCCAGATCAGCATGCCGTCGCCAGTGG 60	
QY	61 GTAATCGCCCTGGCCCTTGGCTGCTGTTGGAGCAGGAGTGGCCAGTGCAGCA 120	
DB	61 GTAATCGCCCTGGCCCTTGGCTGCTGTTGGAGCAGGAGTGGCCAGTGCAGCA 120	
QY	121 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGAGTCTCCACCTGT 180	
DB	121 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGAGTCTCCACCTGT 180	
QY	181 TCCAGATGTCACACCTGTGCTGCGCAGCTGAGTGGCTACATATATCAATGATGCCAG 240	
DB	181 TCCAGATGTCACACCTGTGCTGCGCAGCTGAGTGGCTACATATATCAATGATGCCAG 240	
QY	241 CTCGCTTGGCCCGGATTAACCAAGAGCATTCAGATCATGAAGATGGCAATGC 300	
DB	241 CTCGCTTGGCCCGGATTAACCAAGAGCATTCAGATCATGAAGATGGCAATGC 300	
QY	301 TGATCCACAGAGCAGCTCAAGCCATGATGTCAGTGGAGAACAGTGTGGGCATGG 360	
DB	301 TGATCCACAGAGCAGCTCAAGCCATGATGTCAGTGGAGAACAGTGTGGGCATGG 360	
QY	361 AGAGATATGATGATTAATTAAGATGCCAGCCAACTG 398	

Db 361 AGAGATATGACATGAATTAAGATCCAGCCCACTG 398

RESULT 2

US-08-744-670-2
Sequence 2, Application US/08744670
Patent No. 5858710

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
APPLICANT: Goli, Surya K.

APPLICANT: Murty, Lynn E.

TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/744,670

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0155 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 388 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: SINTTUT01

CLONE: 1539065

US-08-744-670-2

Query Match 97.5%; Score 388; DB 2; Length 388;

Best Local Similarity 100.0%; Pred. No. 8.6e-118;

Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 GCCAGCTCAGGCTACACTATCCAGAGATCGCATGCGCCGTCGCCAGTGGTATGCCCC 70
Db 1 GCCAGCTCAGGCTACACTATCCAGAGATCGCATGCGCCGTCGCCAGTGGTATGCCCC 60
QY 71 TGGCCTTGGCTGCCCTCTCTGTTGTGGACAGGGAAGTCCAGTGGGACAGAGAAAGCTCC 130
Db 61 TGGCCTTGGCTGCCCTCTCTGTTGTGGACAGGGAAGTCCAGTGGGACAGAGAAAGCTCC 120
QY 131 CTTTCTCAAGATGCGCATCTGTGAACACATGCTAGAGTCTCCAACTGTTTCCAGATGT 190
Db 121 CTTTCTCAAGATGCGCATCTGTGAACACATGCTAGAGTCTCCAACTGTTTCCAGATGT 180
QY 191 CCAACCTGCTCTGGGACACTGATGGGCTCACAATATACGAATGAATGCCAGCTCTGCTTGG 250
Db 181 CCAACCTGCTCTGGGACACTGATGGGCTCACAATATACGAATGAATGCCAGCTCTGCTTGG 240
QY 251 CCGGATTAATAAACCAACAGACATCCAGATCATGAAAGATGGCAATGTGATCCACA 310
Db 241 CCGGATTAATAAACCAACAGACATCCAGATCATGAAAGATGGCAATGTGATCCACA 300

QY 311 GGAGCACTTCAAGCCATGAAGTGTCACTGGAGAACAGTGGGCGATGGAGAGATATG 370
Db 301 GGAGCACTTCAAGCCATGAAGTGTCACTGGAGAACAGTGGGCGATGGAGAGATATG 360
QY 371 ACATGAATTAATAAGATCCAGCCCACTG 398
Db 361 ACATGAATTAATAAGATCCAGCCCACTG 388

RESULT 3

US-09-149-933-2

Sequence 2, Application US/09149933

Patent No. 5958699

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Goli, Surya K.

APPLICANT: Murty, Lynn E.

TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/149,933

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0155 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 388 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: SINTTUT01

CLONE: 1539065

US-09-149-933-2

Query Match 97.5%; Score 388; DB 2; Length 388;

Best Local Similarity 100.0%; Pred. No. 8.6e-118;

Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCCAGCTCAGGCTACACTATCCAGAGATCGCATGCGCCGTCGCCAGTGGTATGCCCC 70
Db 1 GCCAGCTCAGGCTACACTATCCAGAGATCGCATGCGCCGTCGCCAGTGGTATGCCCC 60
QY 71 TGGCCTTGGCTGCCCTCTCTGTTGTGGACAGGGAAGTCCAGTGGGACAGAGAAAGCTCC 130
Db 61 TGGCCTTGGCTGCCCTCTCTGTTGTGGACAGGGAAGTCCAGTGGGACAGAGAAAGCTCC 120
QY 131 CTTTCTCAAGATGCGCATCTGTGAACACATGCTAGAGTCTCCAACTGTTTCCAGATGT 190
Db 121 CTTTCTCAAGATGCGCATCTGTGAACACATGCTAGAGTCTCCAACTGTTTCCAGATGT 180
QY 191 CCAACCTGCTCTGGGACACTGATGGGCTCACAATATACGAATGAATGCCAGCTCTGCTTGG 250

Db 181 CCAACCTGCTGCGGCACTATGAGCTACATATAGCAATGAATGCCAGCTCTGTTGG 240
QY 251 CCCGATATATAAACCAAGAGCATCCAGATCATATAAAGATGGCAAAATGCTGATCCACA 310
Db 241 CCCGATATATAAACCAAGAGCATCCAGATCATATAAAGATGGCAAAATGCTGATCCACA 300
QY 311 GGAGCAGCTCAAGCAGATGAGTGTGAGTGGAGAGAGAGAGAGAGAGAGATG 370
Db 301 GGAGCAGCTCAAGCAGATGAGTGTGAGTGGAGAGAGAGAGAGAGAGAGATG 360
QY 371 ACATGAATATAAAGATCCAGCCCAACTG 398
Db 361 ACATGAATATAAAGATCCAGCCCAACTG 388

RESULT 4

US-09-065-019-1

; Sequence 1, Application US/09065019

; Patent No. 6280968

GENERAL INFORMATION:

APPLICANT: Kato, Seishi

APPLICANT: Yamaguchi, Tomoko

APPLICANT: Sekine, Shingo

APPLICANT: Kamata, Kouji

TITLE OF INVENTION: HUMAN PEC-60-LIKE PROTEIN AND DNA ENCODING THIS PROTEIN

FILE REFERENCE: 6700PCT-US

CURRENT APPLICATION NUMBER: US/09/065, 019

CURRENT FILING DATE: 1998-04-17

NUMBER OF SEQ ID NOS: 3

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 180

TYPE: DNA

ORGANISM: Homo sapiens

US-09-065-019-1

Query Match 45.2%; Score 180; DB 3; Length 180;

Best Local Similarity 100.0%; Pred. No. 1.1e-49;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACAGATGTAGATCTCAACCTGT 180
Db 1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACAGATGTAGATCTCAACCTGT 60
QY 181 TCCAGATGTCAACCTGCTGCGGCACTGATGGCTCACATATAGCAATGAATGCCAG 240
Db 61 TCCAGATGTCAACCTGCTGCGGCACTGATGGCTCACATATAGCAATGAATGCCAG 120
QY 241 CTCGCTTGCGCCGCGATATAAACCAAGAGAGATCCAGATCATGAAAGATGCCAATGC 300
Db 121 CTCGCTTGCGCCGCGATATAAACCAAGAGAGATCCAGATCATGAAAGATGCCAATGC 180

RESULT 5

US-08-880-829-21

; Sequence 21, Application US/08880829

; Patent No. 5925559

GENERAL INFORMATION:

APPLICANT: Collins, John

APPLICANT: Roettgen, Peter

TITLE OF INVENTION: A Collection of Phagemids, A

TITLE OF INVENTION: Collection of Escherichia Coli

TITLE OF INVENTION: Cells Carrying The Phagemids, A

TITLE OF INVENTION: Collection of Phagemid Particles

TITLE OF INVENTION: Produced From Said Collection

TITLE OF INVENTION: And Phagemid Particles

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Joseph T. Eisele

ADDRESSEE: Kane, Dalsimer, Sullivan, Kurucz,

ADDRESSEE: Levy, Eisele and Richard

STREET: 711 Third Avenue

STREET: 711 Third Avenue

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017-4059
COMPUTER READABLE FORM:
MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-XT COMPATIBLE
OPERATING SYSTEM: DOS 3.3:
SOFTWARE: WORDPERFECT 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880, 829
FILING DATE: 23-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/458, 668
FILING DATE: 06/02/95
APPLICATION NUMBER: German EP 94 108 689.4
FILING DATE: 06/07/94
ATTORNEY/AGENT INFORMATION:
NAME: EISELE, JOSEPH T.
REGISTRATION NUMBER: 25,331
REFERENCE/DOCKET NUMBER: 2727-77
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 687-6000
TELEFAX: (212) 682-3485
TELEX: (212) 426767
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 base pairs
TYPE: nucleic acid
STRANDEDNESS: single strand
TOPOLOGY: linear
MOLECULE TYPE:
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
US-08-880-829-21

Query Match 10.4%; Score 41.4; DB 2; Length 458;

Best Local Similarity 57.3%; Pred. No. 0.00056;

Matches 75; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 178 TGTCCAGATGTCCAACTGCTGCGGCACTGATGGCTCACATATAGCAATGAATGC 237
Db 220 TGCATAGATGTCCAACTGCTGCGGCACTGATGGCTCACATATAGCAATGAATGC 279
QY 238 CAGCTGCTTGCGCCGCGATATAAACCAAGAGAGATCCAGATCATGAAAGATGCCAAA 297
Db 280 GTTGTGCTTGCGGCAACCGTAACGTCAGACTTCTATCCGATCCAGAAATCTGTCGG 339
QY 298 TGTGATCCCA 308
Db 340 TGTGATCCCA 350

RESULT 6

US-09-016-434-1250

; Sequence 1250, Application US/09016434

; Patent No. 6500938

GENERAL INFORMATION:

APPLICANT: Janice Au-Young

APPLICANT: Jeffrey J. Seilhamer

```
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1250:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g190687
;
US-09-016-434-1250

Query Match
Best Local Similarity 10.2%; Score 40.6; DB 4; Length 432;
Matches 76; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 178 TGTTCAGATGTCACACCTGCTGCGGACATGATGCGCTCACAATATAGATGATGC 237
D 231 TGCACCAAGATATATAGACCTGTCTGTGGAGTGTGAAATCTTATCCCAATGATGC 290
QY 238 CAGCTCTGCTTGGCCCGGATAAACCAACAGACATCCAGATCATGAAAGATGGCAA 297
D 291 GTGTATGTTTAAANAATCGGAACCGCAGACTTCTATCTCATTCAAAATCTGGGCT 350
QY 298 TGCTGATCCACAGG 312
D 351 TGCTGAGAACCAAGG 365

RESULT 7
US-08-235-515A-26
; Sequence 26, Application US/08235515A
; Patent No. 5840518
; GENERAL INFORMATION:
; APPLICANT: Morishita, Hideaki
; APPLICANT: Kanamori, Toshinori
; TITLE OF INVENTION: DNA FRAGMENT, VECTOR CONTAINING THE DNA
; TITLE OF INVENTION: FRAGMENT, TRANSFORMANT TRANSFORMED WITH THE VECTOR AND
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
```

```
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,515A
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1110-135P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8050
; TELEFAX: 703-205-8000
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: 1..302
; LOCATION: 1..302
; OTHER INFORMATION: /label= polynucleotide
; OTHER INFORMATION: /note= "j-x1-y-z(pstl), insert in plasmid pm474,
; OTHER INFORMATION: Figure 5"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 27..89
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 90..293
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..293
;
US-08-235-515A-26

Query Match
Best Local Similarity 9.9%; Score 39.6; DB 2; Length 302;
Matches 72; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 178 TGTTCAGATGTCACACCTGCTGCGGACATGATGCGCTCACAATATAGATGATGC 237
D 171 TGCACCAAGATATATAGACCTGTCTGTGGAGTGTGAAATCTTATCCCAATGATGC 230
QY 238 CAGCTCTGCTTGGCCCGGATAAACCAACAGACATCCAGATCATGAAAGATGGCAA 297
D 231 GTGTATGTTTAAANAATCGGAACCGCAGACTTCTATCTCATTCAAAATCTGGGCT 290
QY 298 TGCTGA 303
D 291 TGCTGA 296

RESULT 8
US-09-221-298-56/C
; Sequence 56, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221,298
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 112
```